

# Example Grant Materials

Do not redistribute except  
under terms noted within.

**Citation:**

Brown, Travis, Jennifer Guiliano, and Trevor Muñoz. "Active OCR: Tightening the Loop in Human Computing for OCR Correction" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities  
Travis Brown  
Paul Evans  
Jennifer Guiliano  
Trevor Muñoz  
Kirsten Keister

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## Active OCR: A Level II Start Up Grant

**Enhancing the humanities through innovation:** Over the past several years, many large archives (such as the [National Library of Australia](#) and the [National Library of Finland](#)) have attempted to improve the quality of their digitized text collections by inviting website visitors to assist with the correction of transcription errors. In the case of print collections, an optical character recognition (OCR) system is typically used to create an initial transcription of the text from scanned page images. While the accuracy of OCR engines such as [Tesseract](#) and [ABBYY FineReader](#) is constantly improving, these systems often perform poorly when confronted with historical typefaces and orthographic conventions. Traditional forms of manual correction are expensive even at a small scale.

Engaging web volunteers—a process often called *crowdsourcing*—is one way for archives to correct their texts at a lower cost and on a larger scale, while also developing a user community. The scope of crowdsourcing is still critically limited, however. For example, the NLA's Australian Newspaper Digitisation Project reports that 2,994 users corrected 104,000 articles during the first six months of the program. While 104,000 articles corrected is substantial, it is proportionally insignificant compared to the amount of text digitized by the [HathiTrust Digital Library](#), for example, which currently holds over 3.3 billion pages, or even to the 182,000 texts that are included in Gale's [Eighteenth-Century Collections Online](#) (ECCO). Additionally, the success of crowdsourcing for OCR correction depends on the amount of public interest in the domain of the collection. A historical newspaper archive, for example, is likely to attract many more committed contributors than a collection of eighteenth-century theological tracts.

*Active learning* is a machine learning approach that attempts to maximize the effectiveness of a human annotator by making the annotation process iterative. Instead of learning a model from a static labeled data set, as in the standard form of supervised learning used in most OCR engines, an active learning system interactively queries the annotator about the instances it finds *most difficult* in some sense. The system prompts the annotator to select an appropriate label and/or provide input as to the example that most closely approximates the one the computer system is attempting to identify. This allows the system to learn more effectively from the human in the “human computing” loop by creating an opportunity for the human to intervene in the data assessment by improving the analytical function of the algorithm itself, not just correcting the individual data point. In traditional crowdsourced transcription applications, a user might spend ten minutes correcting an article, and the product of that labor would be a correct transcription of the article. In an application based on active learning or a similarly iterative approach, on the other hand, the user could identify dozens or hundreds of difficult characters that appear in the articles from that same time period, and the system would use this new knowledge to improve OCR across the entire corpus.

A portion of our funded efforts will focus on the need to incentivize engagement in tasks of this type, whether they are traditionally crowdsourced or engaged through a more active, iterative process like the one we propose. We intend, as we develop an OCR correction system designed to create opportunities for users to improve the system and not just the text itself, to examine how explorations of a users' preferences can improve their engagement with corpora of materials. While our hope is to decrease the number of errors as the algorithm improves,

initial efforts (like most OCR engines) will rely on a large amount of annotator intervention, and potential volunteers will be reluctant to spend significant time correcting errors on documents they are uninterested in. We will explore how analyzing each user's preferences for certain types of documents and suggesting individually tailored queues of relevant materials to annotate may help to create a more engaged and dedicated user community.

We propose a proof-of-concept application that will experiment with the use of active learning and other iterative techniques for the correction of eighteenth-century texts (primarily but not exclusively written in English) provided by the [HathiTrust Digital Library](#) and the 2,231 ECCO text transcriptions released into the public domain by Gale and distributed by the Text Creation Partnership (TCP) and 18thConnect. Through 18thConnect, a scholarly organization focusing on the “long 18th century” in literary scholarship, we will have access to a number of users who would form a potential pool of annotators and evaluators of our proof-of-concept application. We do not see the focus on this period or language as being required for the technical success of the approach, but rather as a way to take advantage of existing resources and collaborations in order to tackle more directly the interesting problems involved in developing these iterative algorithms and the software to implement them. Once the system is developed it could be applied without substantial adaptation to other languages (at least languages supported by the underlying OCR engines) and historical periods. To demonstrate this capability we will test the system on a small multilingual set of early twentieth-century printed materials (including pages digitized by the Whitman Archive) in the final stages of the project.

**Environmental scan:** Our proposed **Active OCR** system would be an experiment in active learning for crowdsourced transcription correction. The effectiveness of such a system is described in Abdulkader and Casey (2009). They argue that although the mean word level error rates for currently functioning OCR rates is between 1 and 10 percent (a seemingly low number), the effect of these rates is prohibitive to information retrieval that has become fundamental to modern research. Using an error estimation algorithm, they run the same document through multiple OCR engines and estimate the error value. The higher the level of error, the greater the need for intervention by human annotators. After clustering the results with those of similar rates, human annotators are asked, from a set of images of individual words taken in context, to select the OCR engine whose algorithm functioned nearest to their own estimation of the word or input their own text. These results are then used to produce the corrected OCR text. We are not aware of any publicly available or demonstrated implementations of this approach—i.e., that uses the input of human annotators to improve the OCR system itself, not just the OCR-transcribed text. None of the existing transcription correction systems that we are aware of (including 18thConnect's [TypeWright](#)) currently use volunteer contributions as training material, and while OCR engines and toolkits such as Tesseract and [Gamera](#) offer OCR training interfaces, these are typically not interactive in the way we describe, in that they require the user to select the examples that are used for training, and they are not collaborative. Furthermore, they do not attempt to identify users' preferences in order to suggest additional annotator interactions.

**History and duration of the project:** MITH has a rich history of research and development on projects involving facsimile images and textual transcriptions, including archives such as the NEH-funded Shakespeare Quarto Archive and tools such as the Text-Image Linking Environment (TILE), a web-based application for creating and editing image-based electronic editions and digital archives of humanities texts (also funded by the NEH). The application we propose will take full advantage of this work and expertise. The project also builds on work done by Travis Brown with Matt Cohen at the University of Texas at Austin on projects for the Walt Whitman Archive. In this work, which Brown and Cohen outlined in a paper presented at the 2011 conference of the Society for Textual Scholarship, Brown used the training capabilities provided by Tesseract along with error estimation techniques to improve OCR output for two large corpora being digitized, transcribed, and encoded by the Whitman Archive. These projects demonstrated the effectiveness of an iterative training-transcription workflow, and the application proposed here will streamline this workflow and make the selection of materials for human consideration more focused and efficient. The project also builds directly on the open-source TypeWright system for collaborative OCR correction developed by [Performant Software](#) for our partners at 18thConnect, and we intend to re-use elements of the TypeWright design and code, in addition to drawing on the expertise of Laura Mandell, the director of 18thConnect. Finally, the focus on the collaborative curation of texts is closely aligned with the goals of [Project Bamboo](#), an international partnership of ten universities, including the University of Maryland. Over the last year MITH has hosted three workshops on the design of corpora-focused functionality in the next phase of Project Bamboo, and the collections and tasks described in this proposal—as well as the TypeWright application specifically—have played an important role in this planning process. For example, the Woodchipper application for text visualization and exploration developed at the first of these workshops operates on HathiTrust and TCP-ECCO texts from the eighteenth and nineteenth centuries, many of which have almost unusable high rates of OCR errors. While that project was focused on exploration, not correction or other forms of curation, we discovered that Latent Dirichlet Allocation topic modeling provided a potentially useful means of characterizing and identifying classes of OCR errors. This kind of interplay between corpus curation and exploration is a core focus of Project Bamboo, and the proof-of-concept application proposed here (and more generally the approach we will explore) could potentially be extended and supported in the future as a part of that project.

**Work plan:** The first stage of development (Months 1-4) will involve the creation of a web-based user interface for editing character box data. A *character box* in this context is simply a pair of coordinates that identifies a rectangular subset of an image containing a single character of the transcription. Different OCR programs store this data in different formats, but these are generally equivalent and easy to map to each other. For example, if the first character on a facsimile page is “T”, the character box file produced by Tesseract may contain a line like the following:

```
T 124 1298 135 1320 0
```

The Tesseract format uses a coordinate system with the origin at the bottom-left corner of the page, so this line indicates that the bounding box containing this character has (124, 1298) as its bottom-left corner and extends to (135, 1320) at its upper right.

There are many open source editors—including several web-based editors, such as a PHP tool provided by Tesseract—that allow character box data files to be edited in a graphical user interface, so that the user can drag box outlines to fit the character in the facsimile image, instead of being required to enter coordinates manually. These tools also generally support common operations such as splitting or merging character boxes (which is necessary, for example, when the OCR system incorrectly recognizes the characters “rn” as “m”). The TypeWright tool also supports word-level bounding box annotation in a collaborative setting.

The development in this first stage will involve selecting and adapting one or more of these existing open source tools that will provide the functionality we need and be compatible with the rest of our software platform. We anticipate that a full-time developer could accomplish this work in approximately two weeks. This work will be shared by Travis Brown (the lead developer) and the research assistant, with the precise division of duties depending on the expertise of the research assistant. All software will be documented by the lead developer and research assistant, and will be released (early and frequently) to the public through a [GitHub repository](#) under Version 2.0 of the Apache License.

The next stage of development (Months 5-7) will involve the creation of several key back-end components, including primarily uniform interfaces to several open source OCR engines that will allow us to gather additional information about their output, such as the value associated with the system’s confidence about each character. In the case of the Tesseract engine, Brown has already done much of this work in projects for the Whitman Archive. Another component will support error estimation, and may be based in part on work done for MITH’s Woodchipper project. These components will potentially operate in tandem, with the following high-level description representing one possible workflow:

1. The error estimator is used to identify the “worst” documents in the sub-corpus of interest to the current user.
2. Multiple OCR engines are used to re-transcribe these particularly erroneous documents, with points of disagreement or low confidence being highlighted as needing prioritization for annotator attention.

Related back-end error analysis services could also be developed during this stage. Because this work involves the development of several components that do not directly depend on each other, it will also be shared between the lead developer and the research assistant, and will be completed by the end of the semester during which the research assistant is employed.

The final stage of development (Months 8-10) will complete the human-computer loop, allowing the corrections of particularly challenging materials by human annotators to be used by the underlying OCR systems as training material. While this is a relatively simple extension to the interface in the case of Tesseract, for example, the challenge will be making this kind of iterative training computationally feasible in a multi-user collaborative environment. We imagine that the largest part of the lead developer’s investment of time will be committed to this stage of development.

The final stage of the project will involve testing and evaluation (Months 10-12). Volunteers will be invited primarily from the [18th Connect](#) community, and all corrections and other data generated by users will be anonymized and made publicly available in a micro-format

similar to that used by Tesseract. Evaluation will be managed by the lead developer and will focus on comparing the transcriptions generated by contributors and the system against traditionally corrected texts, using standard metrics for OCR evaluation.

**Staff:** The proposed project is fortunate to benefit from a variety of substantial relationships at the University of Maryland and MITH. Core project staff will include: Travis Brown, lead Research and Development Software Developer at MITH, who will lead the project and supervise all project activities and Mr. Trevor Muñoz, Assistant Dean of the University of Maryland Libraries and Associate Director of MITH, will be responsible for data management and curation throughout the lifecycle of the project.

**Final product and dissemination:** Our project will release all of its work as open-source code thereby encouraging other researchers to use our work, benefit from our investment of resources, and alter our code to extend the efficacy of Active OCR. We will release a reflective white paper at the end of the grant detailing the challenges and successes of experimenting with machine learning in the digital humanities focusing on OCR. As part of this white paper, we will also gather and archive volunteer contributions on our project website. These, along with archived social media streams like blogs and twitter, will form a useful thread throughout our final reporting efforts as we attempt to extrapolate potential understanding of how active learning can challenge the current approach to OCR. These extrapolations will form a core set of recommended best practices for other developers wishing to join us in further development.

**Citation:**

Dickie, Grant, Jennifer Guiliano, Trevor Muñoz, Jim Smith, and Travis Brown. "ANGLES: A Web-Based XML Editor" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities

Grant Dickie  
Jennifer Guiliano  
Trevor Muñoz  
Jim Smith  
Travis Brown  
Kirsten Keister  
Raffaele Vigilanti

Additional Contributors

Hugh Cayless  
Doug Reside

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

**Citation:**

Dickie, Grant, Jennifer Guiliano, Trevor Muñoz, Jim Smith, and Travis Brown. "ANGLES: A Web-Based XML Editor" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities

Grant Dickie  
Jennifer Guiliano  
Trevor Muñoz  
Jim Smith  
Travis Brown  
Kirsten Keister  
Raffaele Vigilanti

Additional Contributors

Hugh Cayless  
Doug Reside

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## **ANGLES: Level 2 Digital Humanities Start Up**

**Enhancing the humanities through innovation:** Adding machine-processable information—“markup”—to texts, images, or other media is a core activity in many types of digital humanities research. Markup allows scholars to prepare their materials of study for computational analysis, publish these materials in a variety of formats, and record their interpretive judgements in standard form. To fully realize the benefits of working with digital materials in this way, humanities scholars need tools which specifically address their research needs but, at least as critically, humanities scholars need ways of engaging productively with developers and technologists.

The most likely avenue for the development of such tools is the open source software paradigm where “ownership” of technologies is broadly distributed among a community of users and responsibility for innovation and maintenance is undertaken by a distributed set of developers, many of whom are themselves also users. There are many examples of successful projects created using variations of this model, including the Text Encoding Initiative (TEI) standard, the Zotero citation management software, the Fedora Digital Repository software, and the Wordpress publishing platform. The vectors for humanists to engage in or influence the development of these projects are limited—the ability to write code and manage the mechanisms of distributed software collaboration (source code management, ticketing, and similar tools) are common requisites.

As a result, the design phase of tool development in the digital humanities is often either a small group endeavor or a one-to-one exchange where a humanist poses a set of potential research questions and desired functions, a computer scientist or team of developers build the technology, and the humanist returns to provide feedback and then begin using the tool. Driven by a small group of users’ needs, many digital humanities technologies never advance beyond their initial concept—not because the idea was ill-conceived, but because without significant community buy-in, the technology does not fully engage with its full potential user community. Tools are often useful for an individual or a single project but not for the wider community. There should be more points of connection and exchange between the intellectual and methodological activities of humanities scholarship and the development process of software tools intended to serve the digital humanities community. Often, the technical revision process takes into account extensive feedback from user communities only after the technology has been launched and it can be difficult to keep potential users engaged between “releases” of a tool.

We propose a solution to the adoption gap that has grown up between scholars with digital materials and developers of more effective means of conducting research with those materials. By combining the model of intensive code development (a.k.a. the “code sprint”) with testing and feedback by domain experts gathered at nationally recognized disciplinary conferences, we will create a tightly-integrated community-development loop that will allow community buy-in as the tool is being developed. This strategy was used very effectively by UVA for the development of the Hydra digital archives package. To structure interactions between users and coders at these events we will use a test-driven development methods which asks attendees to create “stories” to describe their requirements for the software. This creates “pseudo-code” which the developers can interact with. The project managers will ensure that community input is aligned with the larger development goals and timelines of the project. As a pilot for this “community roadshow” approach to tool development in the humanities, we

will develop a web-based editor for working with XML markup through engagement with the large and active community of scholars, teachers and developers who work with the TEI.

A web-based editor would serve the humanities by providing a lightweight, platform-independent solution that would not require software downloading. Instructors and/or faculty providing XML training would therefore not have to coordinate with IT departments or arrange for the purchase of expensive commercial software. This would dramatically reduce the current resource burden associated with experimenting with XML not just for students but also for humanities faculty interested in leveraging XML for their personal use, regardless of the scale of the project. Additionally, a web-based editor could be more easily integrated into other digital humanities projects since it offers an easy web-based function to markup digital texts. A broad example of this would be digital humanities projects utilizing “crowdsourcing”, the process of asking general users to contribute markup. Having a common web portal for markup would boost buy-in for those kinds of efforts. Even robust digital projects like the Women’s Writers Project at Brown University, the Walt Whitman Archive at the University of Nebraska, and the Music Theatre Online Project at the New York Public Library would potentially be served by this free, web-based XML editor.

Large, successful open source software projects have incorporated techniques such as identifying developer advocates and community managers to coordinate distributed development, encourage conformance with best practices, and provide points of entry for new developers throughout the entire development life-cycle. Zotero, a bibliographic management software package developed by the Roy Rosenzweig Center for History and New Media at George Mason University and Android, the mobile operating system developed by Google, are two notable examples of this process. Combining “code sprints,” tightly-integrated feedback from domain experts, and developer relations—best practices from computer science and open source software engineering—will effectively address the need to generate users for technologies, provide better technical development by potentially having on-demand users to explore code as it is being created, and allow these communities to more clearly elucidate their intellectual, methodological, and technological needs.

**Environmental scan:**The field of XML editors is largely dominated by XML desktop-based editors including <Oxygen/>, Eclipse, and XML Pro. Robust, these tools are quite useful for large-scale projects or major research universities who can invest the financial resources (in the case of <Oxygen/>) and technical resources (in the case of Eclipse and XML Pro.) These tools rarely serve individual scholars or teachers from lower-resource universities and/or the general public. Currently, there are no web-based platform-independent editing tools that treat XML as anything other than plain text.

There are a limited number of platform-dependent tools that leverage XML: 1) The Son of Suda On Line (SoSol) is a web-based, fully audited, version-controlled editing environment being built for the papyrological community using Ruby on Rails (RoR), a programming environment for the Ruby code language. RoR depends on server-side scripts communicating with a centralized database which privileges using SoSol for ancient Greek and Arabic texts. As a result, SoSol restricts itself to a limited range of text markup which then impacts the ability to use the tool online. Users are almost required to use the tool offline which requires an advanced set up procedure in order to accommodate markup associated with prose, printed text, and English manuscripts; 2) The Islandora digital repository software (based on the Drupal content

management system and the Fedora Digital Repository) is currently developing a XML-TEI editing module. In order to implement Islandora software, users must install and configure a Drupal application, then the Fedora Digital Repository, and finally add on the TEI editing module. Users' installation of Drupal can use up memory and resources, making the client browser work slower in addition to a commitment to work within the Drupal and Fedora frameworks; and 3) "Ace: Ajax.org Cloud9 IDE" is a web-based code editor integrated with Github, the popular code management and sharing site. Ace attempts to emulate the environments of lightweight text editors, such as VIM, EMACS, and TextMate. As a general purpose programmer's editor, this tool lacks good support for XML including features such as code/tag completion, schema integration, and xml-aware mechanisms for tracking changes.

Intensive code development for computational tools, while uncommon in the digital humanities community, is frequently used within standard computer science development processes. It allows groups of code developers to rapidly develop a tool (or tool extension) with clear parameters for success within a dedicated and focused development environment. The adaption of this model to digital humanities has been quite successful: the CollateX team utilized a code sprint to contribute to a suite of RESTful services for text collation (2009) and the Center for History and New Media at George Mason University developed Anthologize, a WordPress plugin (2010). We intend to leverage humanists' expertise throughout the complete process and not just at selected moments by attaching the code sprint to professional conferences where humanists can provide immediate feedback as the code is completed.

**History and duration of the project:** ANGLES, constructed during a three-day intensive coding session, was prototyped by Jon Deering of St. Louis University, Hugh Cayless of New York University, and Doug Reside of the New York Public Library. The prototype coding team deployed a commonly-utilized coding process (the code sprint) to rapidly design, develop, and deploy the ANGLES demonstrator supported by remote collaboration tools including Skype and Internet Relay Chat. Working with code sprint attendees including staff at MITH, in its' prototype form, ANGLES provides a simple web-based tool for XML editing that is responsive to the challenges and needs associated with interoperable open-source XML tool usage. It currently provides limited syntax coloring and basic tag completion for a single XML schema. . With this prototype as its basis, the prototype coding team sought additional support from MITH that has resulted in this Level 2 start up application. Our development will be focused on expanding these tools to include customized project schema as well as improving its syntax coloring and tag abilities.

Joining the team for the 1.0 release will be J. Grant Dickie, Trevor Munoz, Jennifer Guiliano, and Jim Smith. Cumulatively, they provide additional expertise in Digital Data Curation, XML-TEI standards, Digital Humanities, and Computational Design that bolsters the already considerable expertise of Deering, Cayless, and Reside. With these additions, the ANGLES 1.0 development team anticipates subsequent phases of this project including: integration with Project Bamboo, a Mellon Foundation-funded project that will provide shared applications and infrastructure for humanities research environments and the integration of Encoded Archival Description schema for archival purposes. As each of these are open-source based community-backed endeavors, we believe that ANGLES 1.0 will potentially see further development from the coding communities embedded in these disciplines including, but not limited to: digital libraries, digital humanities, digital archives, and computer science. The cross-

pollination of our approach of attaching code sprints to professional conferences where specialists in these disciplines will be available for beta-testing, refinement, and feedback is clear. We anticipate these communities taking ownership of the various versions of ANGLES that will be generated leading to our 1.0 release, available as open-source software in Github.

**Work plan:** An open-source integrated development environment (IDE) for editing XML will be developed during the course of three pre-conference code sprints. This IDE will be built upon the work already done by Ajax.org Code Editor (ACE). ACE is an open source, extensible environment built in Python and Javascript programming languages and modular in its architecture. Light-weight in its resource needs and based on the client-side, developers for ACE can expand on the functionality to allow for more writing methods, format and markup support, and integrated tools support. Code sprints will be undertaken by a five-person team composed of Reside, Cayless, Deering, Dickie, and Smith in the two days preceding each conference, where agreed upon features and development "tickets" will be completed. Each sprint will consist of two days of pre-conference coding followed by a day of on-demand coding working with humanists attending the conference. One member of the coding team will be selected to attend each conference and provide on-site integration with developers located remotely. This development will follow a development schedule as outlined below. In brief, this roadmap includes an initial code sprint that will focus on development of IDE "basics", i.e. schema validation, code auto-complete, code coloring and formatting, and backing up data with git. The two remaining code sprints will focus on concerns and suggestions from users, and on development of more customized schema for validation. Please consult the Appendix for an itemized list of activities by conference/code event. Via the chosen developers and social media streams, the complete development team would be actively engaging with scholars as they explore the proposed schema, receiving feedback related to modifications, and coding via three main avenues: 1) through on-site demonstrations (likely during poster sessions or in a book exhibit); 2) through twittering notes related to code modifications and request community feedback; and 3) via potential cross-pollination by requesting assessment by humanists from XML projects attending the event in appointment-style meetings. To aid the development team in managing these interactions and capturing the appropriate social media interactions, a developer manager will be present to facilitate development meetings and assure that features gathered through direct interfacing with conference attendees and Twitter are addressed, analyzed, and made into goals for development. The role of development manager will be shared between Trevor Munoz and Jennifer Guiliano. In this way, the milestones for the project will be acquired directly from crowd-sourcing from users within the field of Digital Humanities and then narrowed down and set into specific coding goals by the lead developer and development manager. We anticipate that we will be able to illustrate the importance of robust integration with tight feedback loops between humanists and computer scientists en masse through this live sprint-conference model.

Development, though, will not cease between these events. The attending developer will collect community feedback and present it to the full development team. While all developers will be required to read and analyze comments from attendees, it will be the job of the lead developer with assistance from the development manager to assess each comment and filter them down to three categories. These categories are: 1.) Necessary Additions or Bug Fixes. These are defined as items that must be included or fixed in the IDE open source code or else

the resulting platform will be utterly useless to the community. This category is reserved for "show-stopper" bugs that either crash or prevent users from completing tasks and functionality that is either already present in similar tools or has a wide-spread demand among the community (i.e. more than a few attendees asked for this feature); 2.) Possible Additions and Quirks. Reserved for those features for which attendees report fewer demands and bugs that can be classified as hard to re-create or are specific anomalies. These items will be solved only if level-1 additions and bugs are already completed; and 3.) Not considered. Reserved only for functionality requests that are either too large to complete in the time allotted the grant or are simply unrealistic. They will not be dealt with. Following this assessment process, these items will be posted under the heading of the conference as disciplinary feedback with the in-progress code being annotated similarly.

A web site will be hosted by MITH that will act as a record of the proceedings of sprints and the comments and interactions from conference attendees. As mentioned, code sprints are difficult to document as they happen quickly and often change directions and goals during the course of a single sprint. Direct recording of such events therefore becomes unmanageable. To accommodate the changing schedule, Github blame records, tickets, and commits will be recorded on the Github repository location and a link provided to these records on the MITH website. These resources will serve as records for posterity as well as ongoing and dynamic records of development for the development team manager to review. After each code sprint, a Github tag, or snapshot of the code, will be set up on the main Github repository. Outside developers and coders will have the choice of "forking the code" (a.k.a. utilizing a segment of code for their own differing track of development) at significant points of development. Reviewers can also view the code and see its progression through the different stages.

Often during start up development projects, the developers are separated from the rest of the test users. This project seeks to experiment with avoiding this facet of the development process. Each developer and manager of the team will be encouraged to have links to their own blogs and incorporate contributions to the main blog on the project website. Biographies and pictures of each of the team members will be presented on a page of the website, with lists of their blog contributions and Github commits and pushes underneath the respective team member. Conference attendees can watch and keep track of development and read blog entries on the website, making the experience of development more interactive and personal. The website can even be designed in a more entertaining and brighter fashion, similar to sites like Github and Twitter. This is done to ensure test users feel comfortable submitting suggestions, bugs, and general feedback to the developers. Likewise, developers will be encouraged to learn from scholars, the research trends in the field as presented in conferences, and each other.

Evaluation of each code sprint will occur after the conclusion of the code sprint and the following conference. Evaluations will focus on what changes were made to the central repository and what milestones were added, removed, or altered during the course of the interactions with conference attendees. The development manager will then evaluate the performance of each coder and go over goals for the team. A larger, all-encompassing evaluation will be conducted following the conclusion of all three code sprints. This will assess the project as a whole and determine next steps for the project. These next steps need to address how the code developed and distributed on Github will continue beyond the code sprints. In addition, documentation will have to be created during the final evaluation for users to

implement the software. This documentation must cover installation of the tool and proper usage of the tool, as well as advanced documentation for developing plugin modules. This documentation will be available on the project website, included with the software, as well as code-specific comments within the source code itself.

**Staff:** ANGLES will be administered by Project PI, J. Grant Dickie, web programmer at MITH. The project team will include: Hugh Cayless (NYU), Jon Deering (St. Louis University), Doug Reside (NYPL), Jennifer Guilliano (MITH), Trevor Munoz (MITH), and Jim Smith (MITH). Additionally, as our project will rely on live on-demand open-source development, we will issue a call for participants on our website and preceding each conference so that additional developers may join the team. They will be selected and managed by the PI.

**Final product and dissemination:** Our project will release all of its work as open-source code thereby encouraging other researchers to use our work, benefit from our investment of resources, and alter our code to extend the efficacy of ANGLES 1.0. We anticipate on-site code sprints and the associated refinements at the 128th Annual Conference of the Modern Language Association (3 to 6 January 2013: Boston, Massachusetts), the Alliance for Digital Humanities Organizations' Annual Digital Humanities Conference (Spring 2013: Lincoln, Nebraska), and Code4Lib 2013 (TBD).

Additionally, we will release a reflective white paper at the end of the grant detailing the challenges and successes of experimenting with the integration of code development via the sprint model embedded within professional conferences. As part of this white paper, we will also gather and archive user interactions with our development team at these events and throughout the development phase via social media apparatus including Twitter and Blogs. These will be archived on our project website and will form a useful thread throughout our final reporting efforts as we attempt to extrapolate potential understanding of how the library, digital humanities, and literature communities wish to interact with XML. These extrapolations will form a core set of recommended best practices for other developers wishing to work in interoperable tool design within these communities.

**Citation:**

Bajcsy, Peter, Kevin Franklin, Karen Fresco, Jennifer Guiliano, Anne D. Hedeman, Rob Kooper, Robert Markley, Tenzing Shaw, and Michael Simeone. "Digging into Image Data to Answer Authorship-Related Questions." National Science Foundation, Grant Submission, University of Illinois, Urbana, Illinois, 2009.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Sheffield, UK  
Humanities Research Institute  
Department of French, School of Modern Languages and Linguistics  
University of Illinois at Urbana-Champaign, USA  
National Center for Supercomputing Applications  
Institute for Computing in Humanities, Arts, and Social Science (I-CHASS)  
Department of French  
Department of English  
Art History, School of Art and Design  
Michigan State University, USA  
MATRIX: Center for Humane Arts, Letters, and Social Sciences  
Department of Computer Science and Engineering  
MSU Museum and Department of Art and Art History  
Alliance of American Quilts, North Carolina, USA

**Team members:**

Peter Ainsworth  
Department of French, School of Modern Languages and Linguistics and Humanities  
Research Institute, University of Sheffield, UK  
Simon Appleford  
Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of  
Illinois at Urbana-Champaign, USA  
Peter Bajcsy  
Image Spatial Data Analysis Group, NCSA, University of Illinois at Urbana-Champaign,  
USA  
Steve Cohen  
MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University,  
USA  
Wayne Dyksen  
MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University,  
USA

Kevin Franklin

Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of Illinois at Urbana-Champaign, USA

Karen Fresco

Department of French, University of Illinois at Urbana-Champaign, USA

Matt Geimer

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Jennifer Guliano

Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of Illinois at Urbana-Champaign, USA

Anne D. Hedeman

Art History, School of Art and Design, University of Illinois at Urbana-Champaign, USA

Anil K. Jain

Department of Computer Science and Engineering, Michigan State University, USA

Rob Kooper

Image Spatial Data Analysis Group, NCSA, University of Illinois at Urbana-Champaign, USA

Mark Kombluh

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Marsha MacDowell

MSU Museum, Department of Art and Art History, Michigan State University, USA

Robert Markley

Department of English, University of Illinois at Urbana-Champaign, USA

Michael Meredith

Department of French, School of Modern Languages and Linguistics, University of Sheffield, UK

Amy Milne

Alliance of American Quilts, North Carolina, USA

Dean Rehberger

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Justine Richardson

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Tenzing Shaw

Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, USA

Michael Simeone

Department of English, University of Illinois at Urbana-Champaign, USA

## **Acknowledgments**

The project is supported by the National Science Foundation and National Endowment for the Humanities from the United Kingdom and the Social Sciences and Humanities Research Council from Canada via a Digging into Data Challenge Grant Award. The material presented is based upon work supported by the National Science Foundation under Grant No. 10-39385.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

## Introduction and Project Aims

For more than 40 years, humanities scholars have used computational analysis to help resolve issues of authorship. Through stylistic and linguistic analysis, researchers have puzzled out answers to questions that range from who wrote *The Federalist Papers* and to who collaborated with Shakespeare on *Henry VIII* and *Pericles*. While determining a writer's "genetic fingerprint" is a difficult task, the wealth of scholarship and algorithms that have developed around printed textual analysis promises to help solve a number of vexing authorship issues as well as expand our knowledge of the written arts. However, in the area of visual arts, computational analysis of authorship has not made the same inroads. To do authorship studies of visual works, scholars must often do painstaking point-by-point analysis of small sets of 2D images of the objects. This work becomes all the more difficult when dealing with cultural artifacts such as quilts, maps and medieval manuscripts that often have corporate and anonymous authors working in community groups, guilds, artisan shops, and scriptoria. Beyond the difficulties of authorship attribution, larger important humanities questions about the influence and migration of artistic elements and patterns become all but impossible to assess when large datasets require individual scholarly inspection of each image. To this end, we propose to address authorship and the corresponding image analyses leading to computationally scalable and accurate data-driven discoveries in image repositories.

This effort will utilize three datasets of visual works -- 15<sup>th</sup>-century manuscripts, 17<sup>th</sup> and 18<sup>th</sup>-century maps, and 19<sup>th</sup> and 20<sup>th</sup>-century quilts. Overarching humanities research questions emerge from these groups of works, such as how visual and production styles reflect regional tastes or historical moments, how traumatic historical events manifest in cultural production, and how artifacts reflect and influence relationships between cultural groups. Together these works present a range of complex authorial relationships and a strong base for advancing knowledge on the research problems inherent to scalable, automated image analyses. Open research problems are divided below into artistic, scientific and technological questions based on the specific datasets that elicit those questions. We expect these questions will be useful across the work of all three groups.

For the 15<sup>th</sup>-century manuscripts, Froissart's Chronicles, the *artistic questions* include: Where and by whom were these manuscripts created? How does a manuscript reflect the tastes of the particular region and historical moment to which it belongs? What does the codicological evidence—scribal hands, catchwords, page layouts, artistic styles in the miniatures and marginal decoration—suggest about book production in this period? The *scientific questions* for Froissart's Chronicles ask: Since these manuscripts were made during the Hundred Years' War, what was the impact of war on culture as measured by the various aspects of these manuscripts, e.g., evidence of patronage? How do they reflect contacts between the cultures of France and England? How do they reflect the ideology of chivalry or the concept of history? The questions for these medieval manuscripts are related to: (a) studying the composition and structure (codicology) of the manuscripts as cultural artifacts of the book trade in later medieval Paris; and (b) identifying the characteristic stylistic, orthographic and iconographic 'signatures' of particular scribes and artists and their collaborators who contributed to the illustration and decoration of the volumes, through the use of image recognition and data mining techniques. A further potential output from identifying scribal hands using image analysis techniques is a process that can transcribe the text from the images, a task that is currently done manually by skilled scholars. Thus not only would the content be subjected to analysis but it might also be possible to process it to allow scholars to perform further text-based mining (although not as part of this proposal) on the previously untouchable textual corpus that is locked away as pixels in an image.

The 17<sup>th</sup>- and 18<sup>th</sup>-century maps come from atlases by Joan Blaeu and Herman Moll (original atlases and digital scans held at the University of Illinois Library). The *artistic questions* for these maps include: What characteristics distinguish individual and corporate groups of artists and engravers? Criteria such as color palette, graphic representations of ships, shading of coastlines, and fonts can be considered as distinctive traits that identify both a) particular artists and engravers, b) the corporate styles developed by the Blaeu family in 17<sup>th</sup>-century Amsterdam (Joan was the son of Willem Blaeu, who founded the largest mapmaking engraving and publishing house in the world) and by Moll and his collaborators who adapted Dutch conventions of mapmaking for English audiences in the early 18<sup>th</sup> century, and c) national styles of depicting specific geographic and manmade features (cities, fortifications, trade centers, etc.). The *scientific and technological questions* are: Do specific maps show a more detailed geographical and/or climatological knowledge in representations of coastlines and harbors? Or navigable rivers? Or shoals and sand bars that pose dangers for ships? Or mountain passes that indicate potential routes for exploration and trade? The scientific and technological questions both influence and are influenced by the artistic

questions. In particular, engravers develop specific artistic techniques for representations that were essential for ships' captains, navigators, and merchants who used published maps to sail often unfamiliar and dangerous waters in South America, Asia, and the Pacific (see Appendix D). Maps therefore negotiate among art, science, trade, and politics, and determining the principles that allows researchers to distinguish among different maps and mapmakers will aid scholars working in the history of science and cartography, art, literary studies, colonial history, and economic history.

For 19<sup>th</sup>- and 20<sup>th</sup>-century quilts, *artistic questions* include: What are the distinct characteristics of an individual quiltmaker or relevant quiltmaking group's choices of pattern selection, fabric and color choices, execution of measurement, layout, needlework and craftsmanship of the pattern design, and, most interestingly, original deviations from traditional patterns? Published quilt patterns became much more common starting in the late 1800s, when certain pattern designers mass-produced their patterns and disseminated them through ladies magazines, and later in syndicated newspaper columns. Geographically dispersed quiltmakers who were exposed to this media began gaining new patterns and pattern ideas. Thus, in a large test bed of documented historic quilts, the societal rise and influence of mass media should be seen through the proliferation of quilts that execute patterns disseminated through syndicated columns. The *scientific questions* include: Can the quilts created by quiltmakers from a cloistered family, community, ethnic, or religious group at a particular time period be differentiated from those of other communities, especially those more exposed to mass media? If so, can changes in the community's participation in mass culture be found through changes in quiltmaking styles? Can a resurgence or interest in a particular historic cultural community's quiltmaking styles be found in quiltmaking a century later? To what extent are quilts made by one Amish family in the 19<sup>th</sup> century similar or dissimilar to those made by urban quilters in the same time period? Does this change over time? Or, from an even more fine-grained perspective, do we find more or less divergence in quilts from the North and from the South? To what extent are quilt patterns regional and to what extent national? Does this change over time? A major theme in American cultural history is the eclipse of regional cultural differences during the 20<sup>th</sup>-century. Can we test that hypothesis by looking at quilts? Can we use the Quilt Index dataset to measure the impact of traumatic historical events—say 9/11 or Pearl Harbor—on American culture? Do we see a measurable change in imagery, colors, or composition after such events?

While identifying distinct characteristics of artists is time-consuming, computer-assisted techniques can help humanists discover salient characteristics and increase the reliability of those findings over a large-volume corpus of digitized images. Computer-assisted techniques can provide an initial bridge from the low-level image units, such as color of pixels, to higher-level semantic concepts such as brush strokes, compositions or quilt patterns. The technological questions are related to the design of algorithms that can extract evidence at the low-level image units that could be aggregated into higher-level semantic concepts and support humanists in image understanding and authorship assignment. The further technological questions are about the statistical confidence of authorship hypotheses obtained by processing volumes of images that could not have been visually inspected with the current human resources within a reasonable time frame. How to extract knowledge about authorship and how to increase our confidence in the characteristics of authorship are the key technological questions.

### Problem Description

Based on the artistic, scientific or technological questions, we formulate and address the problem of finding salient characteristics of artists from two-dimensional (2D) images of historical artifacts. Given a set of 2D images of historical artifacts with known authors, we aim to discover what salient characteristics make an artist different from others, and then to enable statistical learning about individual and collective authorship. The objective of this effort is to learn what is unique about the style of each artist, and to provide the results at a much higher level of confidence than previously has been feasible by exploring a large search space in the semantic gap of image understanding.

### Motivation

Currently, humanists must look at images of historical artifacts to determine distinct characteristics of certain individual (or groups of) miniaturists and map engravers, scribes, quilters, and so on. Such visual inspection involves identifying objects in 2D images, recognizing specific types of objects, discriminating differences among those objects, classifying realizations into groups with similarities, building cumulative evidence over multiple groups of objects with similarity in realization, and assigning a authorship based on temporally evolving expertise in visual inspection. For example, to assign a label of an artistic hand to an illustration in Froissart's Chronicles, we would first

identify objects such as boats, castles, crowns, faces, group of knights, horses, landscapes, skies, spears, tents, towns and water. Next, we would look for and identify the discriminating differences in all found instances of these objects and group the objects based on similarities. Finally, using meticulous book-keeping of all groups with similarities, we would build a mapping between the groups of classified objects and the potential assignment of authorship. This manual process is very labor-intensive and cannot be scaled up to large volumes of digital artifacts. In addition, the salient characteristics (a collection of discriminating differences) per artist are described at a high semantic level, which makes it difficult to automate the discovery process. Thus, there is a need to explore the semantic gap in image understanding and to establish the mappings between pixel level image properties and the higher-level abstract image descriptions.

### Data Repositories and Selection Criteria

*Data repositories:* The primary datasets that we propose to use for this research include:

1. Nine complete Froissart manuscripts from the 15<sup>th</sup> century that have been digitized to similar standards and quality (see Appendix C). These are: Toulouse, Bibliothèque d'Etude et du Patrimoine MS 511, Besançon, Bibliothèque d'Etude et de Conservation MS 864 & MS 865, Stonyhurst College MS 1, Brussels, Bibliothèque Royale MS II 88, MS IV 251 tomes 1 & 2, and Paris, BnF MS S français 2663 and 2664. We are currently seeking funding to add two further complete manuscripts to this dataset: Pierpont Morgan Library MS M.804, and British Library MS Royal 15 E.VI. The current collection of 15<sup>th</sup>-century manuscripts consists of over 6,100 images mainly at 500 DPI, hosted on a federated Storage Resource Broker (SRB) facility between UoS and UIUC using a web-front end collaboratively developed by the two sites (see <http://cbers.shef.ac.uk>). The images can also be retrieved from the SRB system via an API that provides direct access to the image dataset within a programming environment such as the Image To Learn toolset (see technical methodology section).
2. Details on the 17<sup>th</sup>- and 18<sup>th</sup>-century map collections: the University of Illinois Library holds a 1664 Blaeu Atlas and over twenty of the Atlases published by Herman Moll in the early 18<sup>th</sup> century, as well as digital scans of the maps for this project (see Appendix D). These atlases include hundreds of additional maps, and the algorithms developed by this project can be applied to the thousands of pre-1800 maps that are gradually being digitized by libraries across the world. There are currently no systematic means of determining authorship for many of these maps, and the open source software developed by this project will help to encourage more digital scans of these rare and valuable but understudied resources.
3. Details on 19<sup>th</sup>- and 20<sup>th</sup>-century quilt images: the Quilt Index (a partnership of Michigan State University and the Alliance for American Quilts) contains images and detailed information on nearly 25,000 quilts, which will grow to 50,000 by the end of the grant period (see appendix E). The quilts, dating from the 1700s to the present day, are mostly American in origin though the Index will expand to include international collections in the future. Access images (550 pixel-wide JPEG files 72-150 ppi resolution) have been contributed by museums, libraries and documentation projects for education and research use. The set is hosted in MATRIX's open source digital repository, KO RA, and available at [www.quiltindex.org](http://www.quiltindex.org). Many thousands of styles and quilt makers are represented in this dataset as well as a range of image quality depending on original photography. For this project we have selected groupings to address four aspects of authorship: Illinois Amish family quilts from the 1800s, 1930-era Detroit quilts of Mary Schafer (who developed a very distinctive border style), typical turn of the century "crazy" quilts by Iowa quilt maker Lottie Enix, and quilts made by multiple quiltmakers using a published 1930s pattern by artist Eline Foland. Determining salient characteristics of colors, shapes, borders, layouts and patterns with these four distinct groupings will be important to automated clustering which will uncover important new similarities and dissimilarities for a broad range of humanities analysis.

*Selection of data repositories:* Given the overarching goal of understanding characteristics of authorship, the proposed framework should consist of generic image analysis algorithms that could be used or adapted for use on other projects and many other datasets. We have selected datasets that represent three different historical periods and three different media but that raise analogous problems in determining authorship. The purpose of choosing such a variety of datasets is to show how seemingly different humanities research questions can share software and resources. The diversity across the three major datasets permits us to consider the computational scalability and broad applicability of the image analysis algorithms; hence we will not be producing methodologies that are only suitable to one specific type of dataset; they will have a much wider impact and use. The different datasets will further

foster integration and evaluation of algorithms so that common parts across many datasets as well as dataset specific parts of algorithms would be well understood. In addition, work developed across three or more cognate projects is certain to reinforce critical mass and to establish a creative dialogue; solutions that may seem relatively obvious to one project team may prove to be a revelation to another team in the consortium.

### Project Methodology (Approach)

We propose to break down the computing problem of discovering salient characteristics into three low-level semantic components characterizing image content: (1) image representations, (2) feature descriptors, and (3) machine learning methods and similarity metrics for assignments of authorship (Appendix B illustrates a diagrammatic outline of this process).

(1) Image Representation: the image representations refer to various ways in which digital images could represent the information about physical objects. The representations include Color spaces (e.g., RGB, HSV, YUV, CIE) [1], Frequency transforms (e.g., Fourier, Hough or digital cosine transform), Special transforms (e.g., Gabor filters, co-occurrence matrices), Decomposition transforms (principal components, wavelets) and Edge transformations (Sobel, Canny, Robertson, etc. [2]). While there have been studies of what representations are close to human perception following Gestalt psychology principles [3], it has not been established how the image representations map towards discriminating artists and to higher-level semantic descriptions. We plan to explore the search space of the above image representations.

(2) Feature descriptors: Once an image representation has been selected, there is a need to describe properties of image pixels (called features) that capture local and global, deterministic and statistical, spatial and spectral image characteristics. The extraction of features can be specifically designed to focus on color, shape, texture, or motion properties. We plan to explore the search space of the most common features including 1D vector of values, color histogram, shape context [4], or texture descriptors (e.g., extracted from co-occurrence matrices) [5].

(3) Machine learning methods and similarity metrics for assignments of authorship: Given a set of features and classification labels of authorship, there exist many machine learning methods that would generate data-driven models (mappings) to convert input features into a desired set of labels. The data-driven models compare input features using similarity metrics and try to find parameters of a model that would predict the authorship labels accurately. The positive yield of models is evaluated based on the number of correctly classified instances of input features. We plan to explore the search space of multiple machine learning methods including K-nearest neighbors, support vector machine (SVM), artificial neural network (ANN), decision tree, and K-means [6].

Our approach is to explore a large dimensional space consisting of all possible combinations of image representations, feature descriptors, supervised machine learning methods and their parameters in order to select the most salient characteristics per artist. These characteristics per artist are selected based on the accuracy reported by supervised machine learning methods that compare predicted authorship assignment using the data-driven models with the provided authorship labels. The result of such extensive searches would lead to an n-tuple that provides the highest discrimination with reference to two artists. For instance, let us assume that the n-tuple found consists of (a) hue color channel in HSV image representation, (b) frequency of occurrence of each hue value – hue histogram, and (c) similarity of hue histograms measured by chi-squared error and aggregated into groups with similar features using three nearest neighbors. Then, a humanist could interpret the discriminating characteristics of two artists to be a hue component of image colors, a statistical distribution of hue variations in image pixels, and neighboring similarity of hue distributions in the space of all possible statistical distributions. Thus, visual inspections by a humanist would be assisted by a computer-driven recommendation to focus on a hue component of color images and the similarity of hue distributions in images (or the similarity of hue value frequencies across images). This would reduce the search time of a humanist and could change the role of visual inspection from searching to verification and validation. Furthermore, the images would be delivered for visual inspection in the appropriate representation (e.g., hue channel and its hue histogram) rather than leaving a humanist to recover the hue representation from another color space representation by color transforming images inside of his/her brain.

Similarly, these pair-wise (artist-to-artist) analyses would lead to a matrix of discriminating characteristics that could be summarized and presented to a humanist. The summaries provide computer-assisted input into research questions about what salient characteristics of an artist dominate within a group of artists, a school of artists

or a community of artists. Furthermore, they would be tremendously useful in forensic studies when unseen images are presented to determine authorship.

Essentially, and in respect to determining scribal hands, this scientific methodology would help scholars to identifying recurring tell-tale signs pinpointing the work of Scribe A, B or C by providing answers to our scholarly questions such as which forms, or combinations of letters (e.g., ligatures such as 'ct' or 'br'), do our algorithms reveal as being key to distinguishing between A, B and C?

### Computational Requirements

The proposed approach is clearly computationally intensive due to the huge size of the search space for selecting the optimal triplet (the most salient characteristics represented by image representation, features and machine learning methods). The dimensionality can be estimated by multiplying the number of image pairs to be evaluated times the number of image representations times the number of features times the number of machine learning metrics times the number of cross validations. For example, computing the optimal triplet that discriminates between two artists, each represented by 10 images, over 2 color spaces (HSV, H, S, V, HS, HV, SV, RGB, R,G, B, RG, RB, GB), 1 feature type (histogram of each band with the number of bins varying from 100 to 255), one machine learning method (k-nearest neighbors with the k parameter taking values 1, 2,3,4,5) and 5-fold cross validation requires  $(10+1)*5$  of image pairs to be evaluated times 14 image representations times 155 features times 5 machine learning variations times 5 cross validation evaluations. This number is equal to 2,983,750 computations with many floating point operations during accuracy evaluations of machine learning models. Clearly one feature type and one machine learning method will not be sufficient to capture the spectrum of underlying image characteristics that discriminate artists. We anticipate computations to number about 300 million after adding variables on each dimension and reducing the number of parameters based on initial optimizations (e.g., the number of histogram bins).

### Development Methodology & Standards

The Image to Learn (IM2 Learn) toolset<sup>1</sup>, developed at NCSA, provides a suite of image transformation functions that will be applied to our dataset to support the first phase of our investigations. IM2Learn additionally provides an API to facilitate the development of other image transformation algorithms that do not currently exist within the package. This standard API will be used by the project partners while developing further functions to ensure cross-compatibility across the sites.

The feature description catalogue (phase 2) will be recorded using an XML database and document type definition (DTD) that will be developed in consultation between the technical developers and scholars to meet the needs of the project from both perspectives. This will provide a structured document that is both human and machine readable that will be directly input into phase 3.

The machine learning component of our methodology (phase 3) will make use of the open source Weka software<sup>2</sup>, developed by the University of Waikato, which provides a collection of machine learning algorithms and a platform to develop additional functionality to suit this proposal. However, unlike the image transformation phase and the use of IM2Learn, our project may not be best served by a direct uptake of the Weka interface due to an increased complexity in operations, particularly since one of the aims of this initiative is to scale-up the algorithms to work on high-performance computing grids. The technical project team will therefore meet to decide a suitable framework, also taking into account previous decisions from the earlier stages to ensure the process is both streamlined and coherent while the partner sites collaborate on aspects of functionality.

The algorithms contained within the IM2Learn and Weka packages will be enhanced by the inclusion of new and existing code from the project partners converted into the common framework/interface. This activity will be supported by bringing together the algorithms that each partner has developed in past projects which relate to this application. Furthermore, to facilitate scalability, we anticipate that the algorithms that feature in our final pipeline will need to be written and optimized to allow large-scale parallelism. The new algorithms developed within the scope of this proposal will be made available as open source in line with the exit strategy outlined below.

---

<sup>1</sup> <http://isda.ncsa.uiuc.edu/Im2Learn>

<sup>2</sup> <http://www.cs.waikato.ac.nz/ml/weka>

Documentation on the framework and an evaluation of the interfaces, protocols and standards used during the development stages will be published as part of the final report. This will include particular references to the team success in achieving a platform for working collaboratively on interoperable code across geographically remote sites.

### Environmental Scan

Based on our knowledge, there has not been an effort in the past such as the proposed one. There are publications referring to forensic studies of individual photographs of arts investigating the authenticity [7, 8], or the use of optics during art creation [9, 10]. Several researchers have explored certain sub-sets of features, for instance, brush strokes [11, 12], for classification purposes. The closest effort to the proposed one is the work of Shen [13] where 1080 classical paintings from 25 artists were analyzed using a collection of color, texture and shape features. However, the search space in [13] is constrained to only searching over one data representation (CIE) and one machine learning method (neural network). In addition, our proposed effort is different not only in the scale of data sets and required scalability of search computations but also in posing a fundamentally different question of finding salient characteristics discriminating two authors and groups of authors at the low image level by computers and then at the higher semantic level by humanists. This approach has been initially investigated by two coPIs of this project at UIUC (Hedeman and Bajcsy) with funding from the NCSA Fellowship program [14].

### Final Product and Deliverables

The final product will primarily consist of (a) data about salient characteristics of an artist with respect to another artist and with respect to a group of artists, and (b) software for obtaining salient characteristics. The data could be viewed as evidence supporting a authorship assignment based on (1) image-derived primitives including image representation, image features and machine learning model for assigning authorship, and (2) human-defined semantic descriptors of unique characteristics that map into a combination of multiple image-derived primitives.

The final product would be used immediately by scholars for book trade studies, for understanding engravers and cartographic artists, and for addressing questions about collective and individual authorship of quilts. Additional likely users of the final product would be the researchers and students at UIUC, MSU, University of Sheffield, and other universities for educational purposes. We foresee other humanists wanting to know about the authorship of all kinds of collections and using the software framework and the algorithms. The algorithms and software developed will have appropriate technical and user documentation and all content made publicly available as open source for non-commercial use. The final products will demonstrate to a wide audience how large-scale data repositories of cultural heritage materials can change the nature of research for the humanities and social sciences.

### Dissemination

Websites will be created at each of the partner sites at the start of the project to introduce and publicize the work. These sites will act as primary points of contact for project information, offering up-to-date examples of results and blogs (which will be used to inform the final report). The websites will host the final deliverables and include the source code and documentation developed during this initiative.

A workshop in Sheffield is planned to take place around month 9 to bring together other expert scholars in the field of authorship from around the UK. Its primary purpose is to engage with the wider community and understand more broadly how other experts interpret authorship and how the preliminary project results might help their work. We also anticipate that the workshop will influence our choices of image representations during the machine learning phase. The engagement we accomplish during this workshop will also disseminate the work by demonstrating its potential to a wider audience, who themselves have a specific interest in authorship.

Further avenues for dissemination and publication that the UK team will target include the All Hands and Digital Resources for the Humanities and Arts conferences which showcase e-Science tools and methodologies and the Digital Humanities Quarterly journal. Other workshops and conferences will also be attended to disseminate and publish this work. Members of the UIUC team will present status reports on the project at national and international conferences in art history, cultural history, and medieval and eighteenth-century studies.

### Project History

These three partners each have considerable experience with their own datasets represented, and have established working relationships on imaging issues prior to the development of this proposal. In April 2009 members of each partner site were part of an NSF workshop, "Imaging and Image Analyses Applied to Historical Objects," at

the University of Illinois which addressed the process of going from physical historical objects to digital historical objects available via the Internet for educational and research purposes. The overarching theme of the workshop was to understand the challenges associated with imaging and image analysis that are inherent in this process.

The Sheffield project director (PD) is close to completion of the AHRC-funded "Online Froissart" project. The project will deliver an online facsimile/transcription and translation from Books I-III of Froissart's Chronicles, based on digital surrogates of original manuscripts and using XML. The "Online Froissart" will therefore help reinforce mark-up techniques and standards for the present proposal as well as provide knowledge about scribes and artistic hands; research tools emerging from the project, including image-processing algorithms, will also feed into the Digging into Data proposal. The Sheffield PD has also led on the EP SRC-funded "Pegasus" project, which considered grid-enabled interfaces [15] for sharing and displaying in real-time online image datasets and virtual exhibitions to a distributed audience; this was done in partnership with Urbana and used the "Virtual Vellum" software from Sheffield.

With the support of a Faculty Fellowship (2008-09) from NCSA, Co-PIs from the University of Illinois (Hedeman and Bajcsy) began work on developing cyber tools for analyzing the visual imagery embedded in the corpus of Froissart manuscripts (a corpus that will ultimately include the Shrewsbury Book and Morgan MS), in order to provide insights into both the artists' contributions to the construction of these specific books, and more broadly, the functioning of the medieval book trade. The Sheffield and Urbana team have also worked closely on other grant applications including a pending JISC/NEH digitization proposal. NCSA researchers have also explored preliminary automated pattern analysis of "crazy" quilts, which will be part of the dataset and questions addressed in this project. Finally, NCSA and ICHASS have been actively supporting the 18<sup>th</sup> Connect project with computational resources and expertise needed for pre-1800 optical character recognition.

#### Time Plan and Project Management:

- Month 1: Data and initial source code sharing platform established between all partners with all relevant material uploaded. Project meetings will be held over Access Grid. Project websites launched at all three sites. Consortium agreement (JISC requirement) including any new responsibilities, data policies, finalized evaluation plan, etc. arising from first round of project meetings.
- Month 2: Scholars and developers discuss base-line algorithms and selection of training data
- Months 3-4: Integration of existing algorithms and alpha development of new image transformation code from initial conversations. The alpha versions of new code will be rapidly developed (i.e. not optimized) as a means of getting results quickly to determine how useful the algorithm will be long-term.
- Month 5: Scholars determine appropriate feature sets based on transformations with discussions with developers and DTD for XML database decided.
- Month 6: Optimization, modification and validation of image transformation code for large-scale processing.
- Months 7-8: Apply machine learning algorithms to test dataset to determine suitable discriminative n-tuples suitable for authorship classification. Project exposure to classroom students at UIUC, MSU and UoS
- Month 9: Sheffield workshop to discuss results and authorship issues among larger scholarly community. Analyze results to prune misleading representations. Optimize algorithms for large-scale processing.
- Month 10: Re-apply machine learning algorithms to test dataset to determine validity of changes.
- Months 11-12: Final data preparation, computational resource configuration and data processing of complete dataset collection. Presentation of the work at the eScience meeting.
- Months 13-14: Documentation of algorithms and technical processes. Scholarly interpretation of results. Draft final report. Incorporating the project results into teaching materials.
- Month 15: Completion of documentation, final reports (including financial statement) and website updates. Dissemination of project results to classrooms outside of the three partnering universities.

*Note: During development of algorithms, the developers and scholars will continue to work closely together; the milestone months indicated above will be used to consolidate progress; formal meetings will be held to monitor progress across all partners.*

The three project directors from each partner site (Ainsworth, Bajcsy & Rehberger) and project managers (Guiliano & Richardson) will meet formally on a quarterly basis to monitor and report overall progress and respond to unexpected issues that risk causing deviation from the time plan. These meetings will be conducted over access grid/teleconferencing. Each site project director has responsibility for overall management of his/her research teams

(as per funding bodies) and reporting back to the funding councils. The research teams primarily include scholars and technical staff – the project directors also fall into one of these categories – and are responsible for undertaking the work as outlined above.

The project team has been selected across the partner sites to bring together the best cross-disciplinary levels of expertise calculated to deliver the collective goals (see Résumés section for team and individual areas of expertise). Expertise in the different areas will be shared throughout the project to meet the objectives; all technical staff across the sites will work closely together to develop the algorithms, as opposed to each site focusing on a specific dataset. This has the additional benefit of ensuring that work is not duplicated, while cross-site compatibility is assured by the common framework and interfaces that will be used: the Image2Learn API will enforce the standards used to develop the image transformation algorithms; however the machine learning algorithms and optimization steps are more technically involved. Thus to ensure consistency between sites, the Sheffield developer will spend a few weeks at NCSA working with their developers at this stage of the project (approximately month 6/7). During this time, any APIs necessary and methodologies required to run the code across the different sites on different grid infrastructures will be established along with optimization strategies. The coordination of work across the technical staff will be the responsibility of the head technical developer and NCSA project director, Peter Bajcsy. All technical staff will liaise informally via electronic communications.

Dissemination activities will be undertaken by either the research team or specialized personnel depending on how the research teams at each site are organized (see list of participants). Dissemination personnel will be managed by the site PD and overall strategic dissemination policies will be realized across the whole team. Throughout the project all team members will communicate with each other and outcomes will be disseminated throughout the team via the blogs. Formal meeting minutes will be made available on the websites.

As one part of project management, the teams have also discussed budget allocations devoted to the three key components of the project, such as computer science (CS), humanities (H) and dissemination (D). The teams have aligned their resource allocations with the project requirements in order to contribute with complementary expertise. The approximate allocations are (a) JISC funding 80% (CS): 10% (H): 10% (D); NSF funding 73% (CS): 20% (H): 7% (D), and NEH funding 40.1% (CS): 44.6% (H): 15% (D).

### Exit/Sustainability Plan

The project deliverables will be subject to the licensing policies outlined by the Creative Commons initiative, and Illinois open source license<sup>3</sup>, and made publicly available for non-commercial use as open-source, thus fostering adoption within the wider community and use for different datasets. The deliverables will be deposited with JorumOpen (in line with JISC's funding policy) and our project websites for download.

We believe that the scholarly results obtained from the data analysis performed will reach beyond this application with only preliminary findings published in the project lifespan due to the development program and time constraints. The methodologies developed will therefore continue to be used by the team members as new discoveries are found and validated.

From a development perspective, at this stage we do not expect the computer to determine authorship indisputably, but rather to cluster similarities together to allow the scholar to focus on these without trawling through large datasets. The work carried out within this proposal will therefore be used as a platform to bid for further funds from agencies, to allow us and others to continue to explore research directions that this initiative uncovers.

### Risk Management and Intellectual Property Rights

We do not consider there to be any significant risks associated with the proposal. Key staff members are completely committed to the project and we have full rights and permission to use the image datasets in line with the purposes outlined. All existing software we plan to use is open-source. In the event of unexpected changes in staffing, we anticipate that our collective research team and further recruitment will be able to compensate due to the complementary skills that each team member brings to the project.

In terms of the technical validity of this application, a precursor to this work has already been undertaken at NCSA [14], demonstrating promising findings, and thus proof-of-concept, which support the research methodologies set forth in this application. This application will, however, broaden and extend beyond the previous work.

---

<sup>3</sup> <http://www.otm.illinois.edu/node/396>

**Citation:**

Williams, George, Jennifer Guiliano, and Tina Herzberg. "Making the Digital Humanities More Open" National Endowment for the Humanities, Grant Submission, University of South Carolina, Columbia, SC, 2010.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities  
University of South Carolina Upstate

**Team members:**

Maryland Institute for Technology in the Humanities  
Jennifer Guiliano  
James Smith  
Kirsten Keister  
Amanda Visconti

University of South Carolina Upstate  
George Williams  
Tina Herzberg  
Cory Bohon

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## Narrative

### *Enhancing the Humanities through Innovation*

*BrailleSC* combines expertise from the disciplines of English, Education, History, and Computer Science. An accessible, scholarly resource concerning braille literacy in South Carolina, *BrailleSC* (<http://braillesc.org/>) has been online for over a year and has already been visited thousands of times, with individuals from around the world accessing our content. The site includes oral histories from individuals about their experiences with braille in everyday life, pedagogical materials to assist teachers in developing best practices in braille instruction, and resources for families that stress the importance of braille literacy and the methods of braille instruction.

Accessibility is important because disabled users need to be able to participate fully in humanities research and teaching. Further, in providing accessibility tools to disabled communities we are able to enrich their individual research and learning efforts beyond the formal educational process. As the insights of scholars working in disability studies in the humanities have shown, creating tools for individuals with disabilities improves digital environments for all users.

Our project will use NEH funds to continue efforts begun in our Level I Digital Humanities Start-Up Grant to develop accessibility tools that integrate easily with the popular content management and publication platform, WordPress. WordPress, a commonly-used blogging and publishing platform, focuses on aesthetics, web standards, and usability and is available as an open-source download as well. Over the last two years, we have developed plug-ins that have been designed with the needs of visually impaired users in mind. Visually-impaired (VI) users access digital information in different ways than other users. For example, VI users navigate information by listening to a synthesized voice reading textual material aloud to them generated from a software platform referred to as a “screen reader.” To make navigation easier for VI users, our “Access Keys” plug-in allows users to get from page to page and section to section by pressing an easy-to-remember combination of keys. Other users require text enlargement, and our “Text Zoom” plug-in changes the size of the text to meet their needs.

The first phase of our project focused on meeting the needs of blind and low-vision endusers, and we will continue along this successful path by extending the use of Anthologize--which is a free and open source plug-in for WordPress that currently translates RSS text into PDF, ePub, and TEI--to include the conversion of text to contracted braille. As a result, we will not only make it possible to easily convert text into braille files suitable for embossing onto paper, thereby extending humanities content to hundreds of thousands of blind or low-vision readers, but we will also experiment with making braille available visually on-screen through the WordPress interface.

We are requesting \$45,936 in funding to support this effort. Funding to support the development of this tool will set our project apart from most digital humanities projects

by focusing not just on delivering content in a new manner but by directly addressing the needs of under-served, but vital communities. Further, the tools developed as part of this project will enrich any other project utilizing WordPress by providing them with tools to reach a wider audience of endusers.

As a result of our existing collaboration with the South Carolina School for the Deaf and Blind, we have access to users across the state with a variety of sensory disabilities. They will serve as our prototype assessment team and provide an experienced team of assessors with real-life experience.

### ***Technical Innovations Benefiting Humanities Research and Teaching***

With the support of this grant, we will develop a braille translation system as a plug-in for the existing BrailleSC project. Our system will be based on liblouis, the popular open-source braille translation engine <<http://code.google.com/p/liblouis/>>. This engine features support for computer and literary braille, for contracted and uncontracted translation for several dozen natural languages, three braille mathematics and many document formats such as DTBook XML, DocBook, and Microsoft Word XML. Written in C advanced computer language, liblouis can be easily compiled to adapt to variety systems. Our translation tool will be a two-tiered system: the client and the server. The client will be designed as a plug-in to WordPress so that any WordPress-based website can use our system. The client will first transform the online content into an electronic book. The electronic book will be further passed to the server as processing input. Once the result returns, our client will display or print the result in a designated area. The WordPress plug-in Anthologize <<http://www.Anthologize.org>> is already able to transform postings into an electronic book. We will extend Anthologize to include pages and comments. The translation server will exist between the web server and the liblouis translation engine. It will strip requests down to individual commands, pass them to the liblouis translation engine, and return results back to web server. We will write our system in PHP to be compatible with WordPress. (See further details in Appendix.)

### ***Significance and Impact***

***The importance of inclusion:*** Our work aims to increase participation by all people in experiencing and creating scholarly digital projects. The barriers to participation are varied and include such obstacles as the high price of specialized software and hardware, the advanced expertise that such software and hardware often requires, and the thoughtless design choices that can prevent endusers with sensory disabilities from taking full advantage of online resources.

***The importance of proper formatting in braille:*** Correct braille formatting is critically important as it provides clarity for the reader and allows the braille reader to easily navigate the materials. Just as in well-formatted HTML documents, formatting in braille is highly structured and enables readers to quickly pick out information by scanning the page with their hands and jumping from section to section, if necessary. Inconsistent or incorrect formatting can slow the reader and lead to frustration or abandonment of the

document being read. As in print publications, capitalization and punctuation of items in a list would follow the same format in braille as in print. However, lists transcribed into braille are always preceded and followed by a blank line, regardless of whether a blank line would be used in print. In addition, when bullets precede all items in a list in print, then the bullets are ignored in braille. Thus, depending on the format of what is being transcribed, materials may or may not follow the same format as the print version.

***The importance of a free online braille translator:*** Online information is most commonly available to blind individuals via computerized speech output from a screen reading software application. While in some situations, accessing online information with a screen reader may be adequate, listening is not adequate in all situations; especially if an individual needs to have specific information presented in a table or chart (to name just two examples.) Reading braille provides a deeper understanding of the content for many individuals. Yet producing well-formatted braille files is accomplished through one of two expensive methods. First, professionals who are certified braille translators may be hired to create well-formatted braille. Second, a number of commercial braille translation software applications may be used. The most commonly-used applications cost several hundred dollars and are cost-prohibitive to low-income users and non-specialized humanists. Thus, creating a free and easy-to-use online braille translator would make a tremendous difference in the lives of individuals who need braille translations of online content, in the work flow of content creators, and in the ability of humanities researchers to reach braille audiences.

## ***Environmental scan***

### **Braille translation tools**

Currently, a few online braille translation systems are available. These systems include the following:

“libbraille,” which translates Unicode text to braille

<<http://libbraille.org/>>

“Braille Translator” from Mathsisfun

<<http://www.mathsisfun.com/braille-translation.html>>

The “On-Line Braille Generator”

<<http://braille.compelo.com/>>

An “Online Braille Generator,” from Byron Knoll

<<http://people.cs.ubc.ca/~knoll/braille.html>>

These applications allow users to paste snippets of text for translation into braille and then see the result displayed on the screen. This information may not be accessible to individuals using screen readers. Their support for a variety of input formats is limited, and none of them can automatically translate between text and braille.

Additionally, a number of desktop software applications or word exist, such as the following:

Louis braille Translator (free standalone application, Mac-only)

<<http://www.cucacat.org/projects/louis/>>

odt2braille (free plug-in for Open Office Writer, Mac-compatibility problems)

<<http://odt2braille.sourceforge.net/>>

Duxbury braille Translator (\$595.00 for a single user license), and

Duxbury MegaDots (\$595 for a single user license)

<<http://www.duxburysystems.com/>>

These applications feature much more advanced functions than the existing online systems. However, the free applications have fewer advanced features than the commercial applications, and the commercial applications are prohibitively expensive and require training to use in order to produce accurate and well-formatted braille.

Our braille translation tool will be free, user-friendly, and will generate well-formatted contracted braille through an online environment.

### ***History and duration of the project***

Originally funded by a campus-based grant designed to encourage undergraduate research projects, Dr. Williams and Cory Bohon, a USC Upstate Undergraduate student, began exploring the possibilities for making Omeka a more accessible platform in the Fall of 2009. After discussing their progress with Dr. Herzberg in late Fall of 2009, they agreed to collaborate with her on creating an online archive of materials related to braille and braille literacy in everyday life. By the middle of Spring 2010, Mr. Bohon had developed working versions of a few different Omeka plug-ins and a handful of oral histories had been recorded on video and subsequently transcribed. At around the same time, Dr. Williams and Dr. Herzberg approached the Center for Digital Humanities (CDH) in Columbia, SC and were pleased with the enthusiastic response they received to their request for collaboration. Since then, the CDH has provided technical support and server space as needed. The CDH is currently re-configuring their staff and program arrangements, so the current proposal describes a project-specific collaboration between individuals at USC Upstate and those at a different digital humanities center who are also interested in addressing accessibility issues.

Last summer, the accessibility plug-ins were refined. For instance, to make navigation easier for these users, our “Access Keys” plug-in allows users to get from page to page and section to section by pressing an easy-to-remember combination of keys. Additionally, our site underwent extensive usability testing by individuals who have low vision and need screen magnification as well as by blind individuals who use screen readers. Their feedback led to refinements, and we are fully confident that our online resource is totally accessible to individuals who are blind or visually impaired. Their feedback also assisted us in the development of phase two activities and the goals we plan to reach with the support of a Level 2 Start-up grant from the National Endowment

for the Humanities, Office of Digital Humanities. Dr. Jennifer Guiliano, formerly of the Center for Digital Humanities and now with the Maryland Institute for Technology in the Humanities, and her team have conducted a feasibility study, and we are confident that, with adequate funding, we will design and deploy a free and easy-to-use tool that will enable endusers with a variety of disabilities and abilities to access online humanities resources.

### **Work plan, Academic Year 2012-2013**

#### **Month 1**

- Technical Assessment of Louis <<http://www.cucacat.org/projects/louis/>> and Anthologize <<http://anthologize.org/>>; Development of Integration PHP Framework (Williams, Dickie).
- Evaluation of braille produced with Louis (coordinated by Herzberg, Williams)

#### **Month 2-4**

- Programming of Louis to extend it to braille audience (Dickie) via PHP-server design and formulation of processing input

#### **Month 4-5**

- Programming of Louis to extend it to braille audience (Dickie) via PHP-client design and formulation of processing output
- WordPress plug-in front-end display aesthetics design (Keister)
- Evaluation of braille produced with Louis (coordinated by Herzberg, Williams)

#### **Month 6-7**

- Programming of Anthologize to integrate with Louis platform via PHP engine (Dickie)

#### **Month 7-9**

- Integration of Social Media aspects including pages and comments (Dickie)
- Revision of Graphic User Interface Aesthetics based on user evaluation (Keister)

#### **Month 10**

- Creation of Public Documentation regarding Anthologize-braille (Guiliano, Williams, Keister)
- Establishment of Code Repository for all produced code (Dickie)

#### **Month 11-12**

- Braille user testing (Williams, Herzberg)
- Revision of functionality based on user feedback (Dickie)

***Project Staff***

BrailleSC will combine participant expertise from the disciplines of English, education, and computer science. The core members include Dr. George H. Williams, Assistant Professor, Languages, Literature, and Composition department at USC Upstate; Dr. Jennifer Guiliano, Assistant Director, Maryland Institute for Technology in the Humanities at the University of Maryland, College Park; Dr. Tina Herzberg, Assistant Professor, School of Education at USC Upstate; J. Grant Dickie, Programmer, Maryland Institute for Technology in the Humanities at the University of Maryland, College Park; and Kirsten Keister, Graphic Designer, Maryland Institute for Technology in the Humanities at the University of Maryland, College Park.

***Final product and dissemination***

Our code will be released as open-source code and included in both the WordPress plug-in library and GitHub. All content—including oral histories, pedagogical materials, documentation of software, braille files—will continue to be made available under a Creative Commons Attribution-Noncommercial-ShareAlike license. We will develop easy-to-follow instructions for how to use all software we develop. The Center for Digital Humanities at the University of South Carolina, Columbia will continue to provide long-term hosting for all tools and content associated with this project. We will discuss our work at THATCamp CHNM 2013 and will present our work as well as at the 2013 meetings of two conferences: Computers and Writing, and Digital Humanities. Finally, we will release a white paper at the project's conclusion explaining what we've learned about various options for creating online braille documents.

for the Humanities, Office of Digital Humanities. Dr. Jennifer Guiliano, formerly of the Center for Digital Humanities and now with the Maryland Institute for Technology in the Humanities, and her team have conducted a feasibility study, and we are confident that, with adequate funding, we will design and deploy a free and easy-to-use tool that will enable endusers with a variety of disabilities and abilities to access online humanities resources.

### **Work plan, Academic Year 2012-2013**

#### **Month 1**

- Technical Assessment of Louis <<http://www.cucatech.org/projects/louis/>> and Anthologize <<http://anthologize.org/>>; Development of Integration PHP Framework (Williams, Dickie).
- Evaluation of braille produced with Louis (coordinated by Herzberg, Williams)

#### **Month 2-4**

- Programming of Louis to extend it to braille audience (Dickie) via PHP-server design and formulation of processing input

#### **Month 4-5**

- Programming of Louis to extend it to braille audience (Dickie) via PHP-client design and formulation of processing output
- WordPress plug-in front-end display aesthetics design (Keister)
- Evaluation of braille produced with Louis (coordinated by Herzberg, Williams)

#### **Month 6-7**

- Programming of Anthologize to integrate with Louis platform via PHP engine (Dickie)

#### **Month 7-9**

- Integration of Social Media aspects including pages and comments (Dickie)
- Revision of Graphic User Interface Aesthetics based on user evaluation (Keister)

#### **Month 10**

- Creation of Public Documentation regarding Anthologize-braille (Guiliano, Williams, Keister)
- Establishment of Code Repository for all produced code (Dickie)

#### **Month 11-12**

- Braille user testing (Williams, Herzberg)
- Revision of functionality based on user feedback (Dickie)

**Citation:**

Jennifer Guiliano and Trevor Munoz. "Digital Humanities Winter Institute 2013." University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or its personnel.

**Citation:**

Jennifer Guiliano and Trevor Munoz. "Digital Humanities Winter Institute 2013." University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or its personnel.

## ***Digital Humanities Winter Institute 2013: January 7-11, 2013***

Maryland Institute for Technology in the Humanities  
University of Maryland, College Park

### **Institute Summary:**

The Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland proposes to serve as host of the **Digital Humanities Winter Institute**. An extension of the highly successful Digital Humanities Summer Institute (DHSI) at the University of Victoria, DHWI provides a week-long opportunity for intensive learning in skills relevant to practitioners in the digital humanities through coursework, seminars, and lectures. DHWI will take place during the period between the new year's holiday and the start of the spring semester, providing an alternate opportunity for participants who cannot attend the popular Summer Institutes. MITH, a leading digital humanities center that pursues disciplinary innovation and institutional transformation through applied research, is located in College Park, Maryland and acts as a loci for digital humanities efforts in the greater Washington D.C. area. A winter institute in the Washington D.C. area has great potential to expand the reach of existing digital humanities training programs by drawing participants not just from the academic community but also from cultural heritage institutions located in the region including the Folger Library, the Smithsonian, the National Archives, and the Library of Congress. As such, we anticipate roughly 100 participants for the inaugural DHWI with the potential for expansion as interest increases.

### **Host and Sponsors**

DHWI will be hosted by the University of Maryland's Maryland Institute for Technology in the Humanities with support from the University of Maryland Libraries and the College of Arts and Humanities. DHWI has secured additional sponsorships from the Emerging Modernism in Canada Project and the City University of New York (CUNY) [Graduate Center Digital Initiatives Program](#).

### **First Annual Institute Lecture**

Seb Chan, Cooper Hewitt: Sebastian Chan is the former Head of Digital, Social and Emerging Technologies at the [Powerhouse Museum](#). Coming from a background in social policy, journalism and media criticism as well as information technology, he has been building and producing websites and interactive media since the mid 1990s. He is a researcher in several Australian Research Council Linkage projects researching social media, museums, and technology; and is a regular speaker in Australia having presented widely in the cultural sector and also at technology events Web Directions South and CeBIT. He is on the international programme committees of Museums and the Web (USA), Digital Strategies for Heritage (Eu), the Horizon.Au New Media Consortium, and is an International Steering Committee member of Culturemondo, an international group of representatives of cultural portal strategists.

### **DHWI Proposed Courses:**

#### **Core Courses:**

## 1. Project Development

Instructor: [Jennifer Guiliano](#), Assistant Director, [Maryland Institute for Technology in the Humanities](#), University of Maryland

Simon Appleford, Clemson University Cyber Institute

The Project Development course will explore the fundamentals of project design including, but not limited to: formulating appropriate disciplinary questions for digital humanities research, investigating digital humanities tools and resources, structuring your first project, designing publicity and websites for your project, documenting your project work, writing your first grant proposal, and managing your budget.

**2. Humanities Programming** Instructors: [Wayne Graham](#), Head, Research and Development, [Scholars' Lab](#), University of Virginia

[Jeremy Boggs](#), Design Architect, [Scholars' Lab](#), University of Virginia

This course focuses on introducing participants to web development through the use of the Ruby on Rails web application framework. This course will introduce programming and design concepts, project management and planning, workflow, as well as the design, implementation, and deployment of a web-based application. Technologies covered in this course will include git, HTML, CSS, JavaScript, Ruby, Rails, and relational (and non-relational) data stores. Over the course of the week, we will work through the practical implementation of a developing and deploying a small-scale web application.

### Intermediate Courses:

## 3. Data Curation for Digital Humanists:

Instructors: [Trevor Muñoz](#), Associate Director, [Maryland Institute for Technology in the Humanities](#) and Assistant Dean for Digital Humanities Research, University of Maryland Libraries

[Dorothea Salo](#), Faculty Associate, [University of Wisconsin](#)

Data Curation for Digital Humanists: Carrying out computational research with digital materials requires that humanists understand how to manage and curate data over its entire lifetime of interest—not only for their own project but also for reviewers and other scholars seeking to reuse data and publications. Effective data curation involves gaining fluency with publication practices and norms, metadata standards, and the technical characteristics of digital data. Data Curation for Digital Humanists will cover four key areas of data curation: modeling and capturing information for successful curation; understanding infrastructure and systems; navigating legal and ethical issues; and policy, planning and outreach. Participants will understand how to

conduct data curation assessments, write and implement data management plans, better integrate their projects with community standards and technologies, and coordinate with Libraries and IT Groups for access to data curation tools and infrastructure.

#### **4. Exploring Image Analyses**

Instructor: [Michael Meredith](#), Post Doctoral Research Associate, [Humanities Research Institute](#), University of Sheffield, UK

Exploring Image Analysis will engage participants in image analysis techniques for use within their digital research agendas. Participants will be introduced to fundamental concepts in image identification, manipulation, and assessment as well as be given the opportunity for hands on exploration of historical images within digital tools. Participants will be offered the chance to explore their own image dataset, identify existing image algorithms and tools for their use (and/or potential directions to take to build nonexistent tools), and examine what questions can be answered through image analysis.

#### **5. Teaching through Multimedia**

Instructors: [Virginia Kuhn](#), Assistant Professor of Cinema Practice, School of Cinematic Arts, [University of Southern California](#), Associate Director, [Institute for Multimedia Literacy](#) and Director, Honors in Multimedia Scholarship Program at USC

[Cheryl Ball](#), Associate Professor, Department of English, [Illinois State University](#) and editor of [Kairos: A Journal of Rhetoric, Technology, and Pedagogy](#)

Teaching through Multimedia will explore innovative strategies for increasing engagement in the undergraduate and graduate classroom with sound, images, and interactivity. This course explores the interrelationships among technology, culture and communication in order to form a solid foundation for scholarly multimedia authoring. Participants will examine several genres of multimedia scholarship, with the goal of being able to deploy them strategically in a variety of academic contexts.

#### **Advanced Courses:**

#### **6. Large-Scale Text Analysis with R**

Instructor: [Matt Jockers](#), Assistant Professor of Digital Humanities, Department of English, [University of Nebraska](#)

Text collections such as the [Google Books](#) have provided scholars in many fields with convenient access to their materials in digital form, but text analysis at the scale of millions or billions of words still requires the use of tools and methods that may initially seem complex or esoteric to researchers in the humanities. Large-Scale Text Analysis with [R](#) will provide a practical introduction to a range of text analysis tools and methods. The course will include units on data extraction, stylistic analysis, authorship attribution, genre detection, gender detection, unsupervised clustering, supervised classification, topic modeling, and sentiment analysis. The main computing environment for the course will be R, "the open source programming language and software environment for statistical computing and graphics." While no programming

experience is required, students should have basic computer skills and be familiar with their computer's file system and comfortable with the command line. The course will cover best practices in data gathering and preparation, as well as addressing some of the theoretical questions that arise when employing a quantitative methodology for the study of literature. Participants will be given a "sample corpus" to use in class exercises, but some class time will be available for independent work and participants are encouraged to bring their own text corpora and research questions so they may apply their newly learned skills to projects of their own.

## 7. Publishing and Using Linked Open Data

Instructor: [Richard Urban](#), Visiting Assistant Professor, [Florida State University](#)

The publication of structured knowledge representations and open data on the Web opens new possibilities for collaboration among humanities researchers and cultural heritage organizations. This course will introduce participants to the core principles of Linked Open Data (LOD), techniques for building and understanding LOD models, how to locate LOD sources for research, tools for manipulating, visualizing, and integrating available data, and best practice methodologies for publicizing and sharing datasets.

## 8. Digital Editions

Instructor: [Dean Irvine](#), Associate Professor, Department of English, [Dalhousie University](#)

This course is designed for individuals and groups who are interested in creating scholarly digital editions. Topics covered will include an overview of planning and project management, workflow and labour issues, and tools available for edition production. We will be working with the Modernist Commons (<http://modernistcommons.ca>), a collaborative digital editing environment and repository designed by the Editing Modernism in Canada (EMiC) project in collaboration with Islandora and its software-services company DiscoveryGarden. We will work on both text- and image-based editions, following a modularized edition-production workflow--from ingesting images, processing texts with optical-character-recognition software, uploading born-digital content, performing markup on transcriptions and images, collating variant texts, and displaying text and apparatus in different viewers. By the end of the course, participants will have worked through the practical implementation of a modular, small-scale edition prototype. Basic knowledge of TEI and some familiarity with RDF (specifically the standards of the Open Annotation Collaboration) is strongly recommended but not required. The seminar is open to everyone, although it is specifically tailored to participants of the EMiC project. Participants need not be modernists or Canadianists to take advantage of using open-source software and learning best practices for scholarly editing in digital media.

**Anticipated Attendance:** DHWI courses will be set at a maximum enrollment of 20 students per courses. All courses will be open-enrollment with at least 5 spots reserved in each course for scholarship students and sponsored attendees.

**Additional Activities:**

### **Graduate Student Networking Event**

Hosted by: the City University of New York (CUNY) Graduate Center Digital Initiatives Program  
The Graduate Student Networking Event, open to all DHWI graduate students, will allow young scholars to engage with one another, discuss issues of concern in DH, and speak with leading digital humanists and instructors. Refreshments will be provided.

### **API Workshop**

Hosted by:

[Dave Lester](#), Creative Lead, [Maryland Institute for Technology in the Humanities](#)

[Ed Summers](#), Information Technology Specialist, [Library of Congress](#)

The API workshop will provide attendees with the hands-on opportunity to learn about APIs for use with cultural heritage datasets to improve digital access to cultural heritage collections. This event will be limited to the first 25 attendees, and is open to participants at an introductory level.

### **The Million Syllabi Project Hack-a-thon**

Hosted by: [Dave Lester](#), Creative Lead, [Maryland Institute for Technology in the Humanities](#)

The Million Syllabi Project Hack-a-thon will explore new ways of using [the million syllabi dataset](#) gathered by Dan Cohen's Syllabus Finder tool from 2002 to 2009. Join a group of hackers exploring teaching, pedagogy, and the syllabus.

### **[Humanities, Plain & Simple](#)**

Institute participants will be asked to contribute to the 4Humanities campaign to document how Humanities-based thinking has directly or indirectly altered or innovated their strategies, ideas, research, leadership, and learning. Participants are invited to: -record short 30 second statement in the importance of the humanities in their lives -contribute a multimedia piece -post a blog -submit a code snippet

### **DHWI Ignite**

DHWI Ignite showcases new and emerging, innovative and engaging work of those at DHWI by providing the opportunity for 5-minute stand up talks during the opening reception the first night of the Institute. Open to Institute attendees and digital humanists in the DC region, DHWI Ignite is a social networking event where presenters give 5-minute fun, thought-provoking, presentations on topics of their choice in the digital humanities. A Call for Ignite Talks will be distributed in late summer 2012.

### **Proposed Schedule of Events (subject to change):**

Sunday, January 6, 2013

6:00-7:30 pm

DHWI Registration: Marriott Lobby

7:30 pm

Instructors and Speaker Dinner

Monday, January 7, 2013

8:30-9:00 am	Registration
9:00-9:30 am	Welcome
9:30- 10:45	Class Session 1
10:45-11:00 am	Break
11:00 am- Noon	Class Session 2
Noon- 1 pm	Lunch (provided)
1:00-2:30 pm	Class Session 3
2:30-2:45 pm	Break (snacks provided)
2:45-4:00 pm	Class Session 4
4:00-5:00 pm	Lecture: Speaker TBA
5:30 pm-7:00 pm	Graduate Student Networking Event
7:00 pm	Dinner (on your own)

Tuesday, January 8, 2013

8:30-9:00 am	Continental Breakfast (provided)
9:00 am- 10:30 am	Class Session 5
10:30-10:45 am	break
10:45- Noon	Class Session 6
Noon- 1:00 pm	Lunch (provided)
1:00- 2:45 pm	Class Session 7
2:45- 3:00 pm	break (snacks provided)
3:00-5:00 pm	Class Session 8
5:30-7:00 pm	DHWI Public DH: API Workshop
7:00 pm	Dinner on your own

Wednesday, January 9, 2013

8:30-9:00 am	Continental Breakfast (provided)
9:00-10:00 am	Lecture: Speaker T.B.A.
10:00-10:15 am	Break
10:15-Noon	Class Session 9
Noon- 1:00 pm	Lunch (provided)
1:00-2:45 pm	Class Session 10
2:45-3 pm	Break (snacks provided)
2:45-5:30	DHWI Ignite
5:30-8:30	DHWI Public DH: Millions Syllabi Project Hackathon

Thursday, January 10, 2013

8:30-9:00 am	Continental Breakfast (provided)
9-10:30 am	Class Session 11
10:30-10:45 am	break
10:45-Noon	Class Session 12
Noon- 1:00 pm	Lunch (provided)
1:00-3:00 pm	Class Session 13
3:00-3:15 pm	Break (snacks provided)
3:15-5:00 pm	Class Session 14
5:30-7:30 pm	DHWI Banquet Humanities Plain and Simple

Friday, January 11, 2013

8:30-9:00 am	Continental Breakfast (provided)
9:00-10:45 am	Class Session 15

10:45 am- 11:00	break
11:00-Noon	Class Session 16
Noon- 1:00 pm	Lunch (provided)
1:00-2:30 pm	Class Wrap-Up
2:30-2:45 pm	Break (snacks provided)
2:45-3:45 pm	Show and Tell
3:45-4:30 pm	Individual Consultations and Partnership Meetings (optional)

**Citation:**

Guiliano, Jennifer, and Kevin Franklin. "Humanities High Performance Computing Collaboratory" National Endowment for the Humanities, Grant Submission, University of Illinois, Urbana, IL, 2008.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Illinois  
Institute for Computing in Humanities, Arts, and Social Science  
Pittsburgh Supercoming Center  
San Diego Supercomputer Center  
The Scholarly Community for the Globalization of the "Middle Ages" (SCGMA) Group  
The Humanistic Algorithms: Semantic Networks in Multimedia Scholarship Group  
The HistorySpace Project: Information Rich Virtual Environments for Historical Scholarship Group

**Team members:**

University of Illinois  
Simon Appleford  
Jennifer Guiliano  
Alan Craig  
Alex Yahja  
Kevin Franklin  
Vernon Burton  
Peter Bajcsy  
Jim Onderdonk  
Pittsburgh Supercoming Center  
Laura McGinnis  
Shawn Brown  
Tom Maiden  
John Urbanic  
San Diego Supercomputer Center  
Diane Baxter  
Reagan Moore  
Richard Marciano

Allsen, Thomas  
Asher, Catherine B.

Aytes, Ayhan

Boone, Jim

Professor Emeritus, College of New Jersey  
Associate Professor of Art History, University of Minnesota  
Visual Media Researcher and Graduate Student in Communications, University of California- San Diego  
Associate Professor of Anthropology, University of New Mexico

Goldberg, David Theo	Director, University of California Humanities Research Institute (UCHRI); Professor of Comparative Literature and Criminology, Law, and Society at the University of California, Irvine.
Hart, Roger	Assistant Professor of History, University of Texas at Austin.
Heng, Geraldine	Director of Medieval Studies, Associate Professor of English; Holder of the Perceval Endowment in Medieval Romance, Historiography, and Culture at the University of Texas at Austin.
Ilmitchi, Gabriela	Assistant Professor of Musicology, University of Minnesota.
Kea, Ray A.	Professor of History, University of California, Riverside
Klieman, Kairn A.	Associate Professor of History, University of Houston, Texas.
Larkin, Margaret.	Professor of Arabic Literature, Department of Near Eastern Studies, University of California, Berkeley.
Liu, Xinru	Assistant Professor of Early Indian History and World History, College of New Jersey.
Martin, Will	Graduate Student, Department of Information Science, University of Texas, Austin.
McIntosh, Susan K.	Professor of Archaeology, Rice University.
Noakes, Susan	Director, Center for Medieval Studies, Professor of French and Italian, University of Minnesota, Twin Cities.
Phillips Jr., William D.	Director, Center for Early Modern History, Professor of History, University of Minnesota
Ragnow, Marguerite	Curator, James Ford Bell Library, University of Minnesota.
Saldanha, Arun	Assistant Professor, Geography, University of Minnesota.
Schmidt, Peter R.	Professor of Anthropology, University of Florida
Spellberg, Denise	Associate Professor of History and Middle Eastern Studies, Associate Director of Medieval Studies, University of Texas.
Urbam, Tomislav	Data and Information Service Group, Texas Advanced Computing Center.
Waltner, Ann B.	Director, Institute for Advanced Study, University of Minnesota.
Widner, Mike	Graduate Student, Medieval Literature, University of Texas, Austin.
Wilcox, Rebecca	Graduate Student
<b>The Humanistic Algorithms: Semantic Networks in Multimedia Scholarship Group:</b>	
Arroyo, Sarah J.	Assistant Professor of English, California State University at Long Beach
Ball, Cheryl	Assistant Professor, English, Illinois State University; Editor of Kairos, refereed online journal exploring the intersections of rhetoric,

technology, and pedagogy

Blalock, Glenn	Assistant Professor of Composition and Rhetoric, Department of English, Baylor University; Creator of CompPile, a scholarly database of journals in rhetoric and compositions.
Callahan, Vicki	Founder of the Conceptual Studies in Media Arts Production at the Peck School of the Arts (Film Dept), University of Wisconsin, Milwaukee
Franklin, Kevin	Executive Director, Institute for Computing in the Humanities, Art, and Social Science, University of Illinois
Haswell, Rich	Professor Emeritus, Department of English, Texas A & M, Corpus Christi; Creator of CompPile, a scholarly database of journals in rhetoric and compositions.
Hawisher, Gail	Professor of English and Director of the Center for Writing Studies, University of Illinois.
Holmes-Wong, Deborah	Archivist for Digital Media, University of Southern California
Kuhn, Virginia	Associate Director, Institute for Multimedia Literacy, University of Southern California
Selfe, Cynthia L.	Humanities Distinguished Professor, Department of English, Ohio State University
Vitanza, Victor J.	Professor of English and Director of the Ph.D. Program in Rhetorics, Communication, and Information Design, Clemson University
Wright, Elijah	Adjunct Lecturer and Graduate Student, School of Library and Information Sciences, Indiana University
Wysocki, Anne F.	Associate Professor English, University of Wisconsin, Milwaukee
The HistorySpace Project: Information Rich Virtual Environments for Historical Scholarship Group: Ayers, Edward	President, University of Richmond; Aurora Project; Valley of the Shadow Project.
Bonnett, John	Assistant Professor, Humanities Institute, Department of History, Brock University (Canada)
Denard, Hugh	Professor, Centre for Computing in the Humanities, Director, King's Visualization, Lab King's College London; Editor, Didaskalia; Member, The Pompey Project
Dunae, Patrick	Professor, Department of History, University of Victoria; Co-Director, Vi.History.ca (online); editor of The Homeroom, British Columbia's online history of education website.
Dunn, Stuart	Research Associate, Centre for Computing in the Humanities, King's College London.
Gilliland, Jason	Assistant Professor, Department of Geography; Director, Urban Development Program,

Lutz, John

University of Western Ontario.  
Professor, Department of History, University of  
Victoria; Co-Director, Great Unsolved Mysteries  
in Canadian History (online); Co-Director,  
Victoria's Victoria (online); Co-Director,  
Vi.History.ca (online); Co-Director, Who Killed  
William Robinson? Race Justice and Settling the  
Land (online).

Thomas III, William G.

John and Catherine Angle Professor in the  
Humanities Department of History, University of  
Nebraska; Aurora Project

### **Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## **Humanities High Performance Computing Collaboratory (H<sup>P</sup>C): “Coordinating High Performance Computing Institutes and the Digital Humanities”**

### **Description of Project and Its Significance**

The Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois at Urbana-Champaign will lead a collaboration partnering the National Center for Supercomputing Applications (NCSA), the Pittsburgh Supercomputer Center (PSC), and the San Diego Supercomputer Center (SDSC) that will foster innovation in the research and development of computational resources for humanities research groups. ***Humanities High Performance Computing Collaboratory (H<sup>P</sup>C): Coordinating High Performance Computing Institutes and the Digital Humanities*** will engage scholars in sustained collaboration with high performance computing specialists in order to identify, create, and adapt computational tools and methods. “Humanities High Performance Computing” signals an investment (structural, computational, and resource-based) in the extension of H<sup>P</sup>C to serve as a portal for humanities scholars with all levels of expertise—from beginner to the most advanced—to receive technical support, access to high performance computing, and products and services associated with the digital technologies. Participants will consult with each computing staff about digital technology—Simulation and Modeling, Social Networking, Grid and Distributed Computing, Data Analytics, or Visualization technologies, discuss these technologies via a virtual community, and develop long-term technological goals for their projects (Appendix II: Collaboration Diagram). H<sup>P</sup>C will facilitate nine mini-residencies (three per center) as well as a two-day conference for forty-five humanities scholars and fifteen high performance computing specialists. In addition, the grant will support the construction and maintenance of a virtual community for participants and the larger public which will function as an online collaboratory space.

## Table of Contents

Description of Project and Its Significance.....	1
Table of Contents.....	2
Significance.....	3
Institutional Profile.....	4
Curriculum and Work Plan.....	5
Participants.....	6
Impact and Evaluation.....	8
Staff, Faculty, and Consultants.....	9
Budget Justification and Budget.....	11
Appendix I: References.....	18
Appendix II: Collaboration Flowchart.....	20
Appendix III: List of Participants By Group.....	21
Appendix IV: Curriculum Workplan Itinerary.....	24
Appendix V: Curriculum Vitae- I-CHASS Faculty, Staff, and Consultants.....	26
Appendix VI: Curriculum Vitae- Humanities Group Leaders.....	38
Appendix VII: Curriculum Vitae- High Performance Computing Center Staffs.....	60
Appendix VIII: Previous Workshop Experiences.....	67
Appendix IX: Computational and Logistical Resources.....	68
Appendix X: Letters of Commitment and Support.....	71

## Significance:

Over the past twenty years, the interdisciplinary field of humanities computing has created a riot of research information and tools, developed by many institutions across a variety of technical formats and platforms that are often unfamiliar to humanities scholars who rely on personal computing and not high performance computing.<sup>1</sup> The proposed *Humanities High Performance Computing<sup>2</sup> Collaboratory (H<sup>P</sup>C): Coordinating High Performance Computing Institutes and the Digital Humanities* project will address this gap between software, computing hardware, and humanities research by creating a collaborative structure that will serve as a consulting and development resource for advanced humanities scholars. Bridging the divide between existing technologies, supercomputing centers, and the innovative work of humanities scholars, the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois at Urbana-Champaign will lead a collaboration<sup>3</sup> partnering the National Center for Supercomputing Applications (NCSA), the Pittsburgh Supercomputer Center (PSC), and the San Diego Supercomputer Center (SDSC) that will foster innovation in the research and development of the comprehensive information technology infrastructure, termed “cyberinfrastructure,” for selected humanities research groups. H<sup>P</sup>C will engage humanists in sustained dialogue with colleagues at NCSA, PSC, and SDSC, in order to identify, create, and adapt computational tools and methods that will enhance and accelerate humanities research and education. Humanities scholars will consult with three separate high performance computing staffs about different technologies, facilitate discussion about these technologies among humanities research groups via a virtual community, and develop long-term technological goals for each project and the digital humanities more generally. I-CHASS will serve as the coordination point for all H<sup>P</sup>C participants. This grant will facilitate nine mini-residencies (three per center) and a two-day conference for forty-five humanities participants from outside of each center’s region, as well as fifteen high performance computing specialists at. In addition, the grant will support the construction and maintenance of a virtual community for participants and the larger public which will function as an online collaboratory space.

This high performance computing-humanities partnership will integrate hardware for computing, data and networks, digitally-enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools with the specialized needs of humanities research groups. This grant will help realize the power of cyberinfrastructure to create, disseminate, and preserve data, information, and knowledge across multiple centers and humanities research groups. Effective technological developments undertaken in this grant will allow humanities scholars to focus their intellectual and scholarly energies on the issues that engage them while simultaneously training them to be effective users of new media and technologies. We envision that each collaborating center’s mini-residence will evolve with new technologies and methods as time passes and that each year a different set of advanced humanities research groups will participate and take advantage of the infrastructure and results generated by this grant. Historian Edward L. Ayers conceived of such partnership in 1991 when became clear that hypermedia and high performance computing offered new possibilities for doing local studies. Ayers commented that much of the work of developing the award-winning *Valley of the Shadow*<sup>4</sup> project was analogous to building a printing press when none existed. Seventeen years later, projects like this are becoming more common within the humanities; yet the vast majority of humanities scholars are still daunted by the humanities-computational sciences divide. They do not have the computational expertise necessary to exploit high performance computing software and hardware to support their research. H<sup>P</sup>Cs’ collaboration will bring together advanced topics in humanities research with high

performance computing staffs to train and tailor each technology to the needs of individual humanities groups. It will build on the experiences and successes of seven humanities-high performance computing programs held at locations including NCSA and SDSC with members of I-CHASS and the Software Environment for the Advancement of Scholarly Research (SEASR) group at NCSA. (See Appendix VI: Previous Workshop Experience.) While the summer workshops provided a broad-based introduction to digital tools for the humanities, arts and social sciences, H<sup>P</sup>C will foster intensive collaboration between humanities research groups and multiple high performance computing centers that result in specific technical and human infrastructures to support humanities research. H<sup>P</sup>C will consist of a core set of presentations and hands-on sessions in computational technologies for humanities activities and will move beyond the simple introduction of available resources and expertise of each collaborating high performance computing center to map these computational resources to the needs of each humanities research group. More simply, invited groups will be paired with high performance computing experts in a particular subfield (Simulation and Modeling, Social Networking, Grid and Distributed Computing, Data Analytics, and Visualization) who will work with the group to introduce, create, and adapt existing technology for use within their project. By the close of H<sup>P</sup>C, each humanities group will have a clearly delineated pathway for the inclusion of specific technologies into their work and research and will have begun to adapt these resources to meet their particular needs.

### **Institutional Profile:**

The University of Illinois at Urbana-Champaign has long been at the forefront of science, technology, and burgeoning computer power beginning fifteen years ago when MOSAIC, the world's first web browser, was developed at the National Center for Supercomputing Applications (NCSA). MOSAIC has revolutionized research and teaching because it democratized access to information. Now bringing the same advances to the humanities, I-CHASS charts new ground in high performance computing and the humanities, arts, and social sciences by creating both learning environments and spaces for digital discovery. Founded in 2004, I-CHASS presents path-breaking research, computational resources, collaborative tools, and educational programming to showcase the future of the humanities, arts, and social sciences by engaging visionary scholars from across the globe to demonstrate approaches that interface advanced interdisciplinary research with high-performance computing. I-CHASS provides these researchers with leadership-class computational resources, both human and technical, to enhance their knowledge discovery and exploration. I-CHASS maintains strategic partnerships with the National Center for Supercomputing Applications (NCSA), the Great Lakes Consortium for Petascale Computation (GLCPC)<sup>5</sup>, and the Illinois Informatics Institute (I3)<sup>6</sup>. These alliances uniquely position I-CHASS to offer unprecedented intellectual and technical expertise to aid humanities researchers in producing interdisciplinary research solutions that will positively affect the human condition and experience.

*Organizational Structure:* I-CHASS currently maintains a core staff of seven employees who work closely with humanities faculty, computing researchers, and research groups from across the globe. The Director of I-CHASS, Professor Vernon Burton, reports directly to the University of Illinois' Provost and meets weekly with the Executive Director, Dr. Kevin Franklin. The Executive Director ensures the day-to-day running of the Institute's operations and consults regularly with the Director, the I-CHASS Advisory Board, and NCSA leaders to discuss issues surrounding the Institute's strategic direction. The I-CHASS staff (Appendix V: Curriculum Vitae- I-CHASS Faculty and Staff) is complemented by the considerable experience and expertise of the 250 plus researchers, technology specialists, and staff who work at NCSA and from whom I-CHASS is able to draw upon as opportunities arise.

*Facilities and Resources:* Founded in 2004 as the only Digital Humanities Center birthed and hosted by a national supercomputer center, I-CHASS is uniquely positioned to strengthen digital humanities scholarship. I-CHASS is housed at NCSA and its Research I institution, the University of Illinois. NCSA is invaluable by bringing massive computational resources (Appendix VII: Computational and Logistical Resources) to the humanities by enabling new and more complex projects nationwide. NCSA employs top-level experts in every field crucial to humanities computing: sophisticated search and retrieval, data management and visualization, human-computer interaction, distributed, collaborative computing, and large-scale modeling and simulation. The computing expertise and resources available at NCSA, harnessed and channeled for the humanities through I-CHASS, can create cyberenvironments for digital-humanities research and entire virtual worlds for interactive education and study in the humanities. Further, as partners in TeraGrid, an open scientific discovery infrastructure combining leadership class resources at eleven partner sites to create an integrated, persistent computational resource, NCSA, PSC, and SDSC, have resources that include more than 750 teraflops of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Researchers can also access more than 100 discipline-specific databases.

*Budget:* H<sup>P</sup>C is requesting \$XXX to fund nine mini-residencies, a two-day conference, and an online collaborative community that will serve forty-five humanities participants from outside of each center's region. This is in addition to fifteen high performance computing center staff participants. Cumulatively, then this grant will serve sixty individuals, not including virtual members who will benefit from the online community and twenty-one participants from humanities institutions local to each center's area (seven per center).

### **Curriculum Overview:**

There will be nine two-day mini-residencies held by technical staffs (Appendix VII: Curriculum Vitae-High Performance Computing Center Staffs) at PSC, SDSC and NCSA. The humanities groups will each rotate through one supercomputer center per quarter- three center-driven residencies for a total of six days for this grant. The centers, working collaboratively, will each focus on training and tool building in their specific area of cyberinfrastructure excellence (Appendix IV: Curriculum Work Plan Itinerary and Appendix II: Collaboration Flowchart). The humanities research groups will receive a broad range of technical and human resource support for multiple technologies that will reinforce each other within their research agenda. The two-day culminating conference will bring together all participants to demonstrate prototypes of technology and tools that have been created/adapted, to disseminate their research experiences across the mini-residences, and to formulate a working paper "Coordinating High Performance Computing Institutes and Digital Humanities" which will chart long-term planning goals to cement these partnerships and further humanities research and computing technologies.

The Pittsburgh Supercomputing Center (PSC)<sup>7</sup> is a joint effort of Carnegie Mellon University and the University of Pittsburgh together with Westinghouse Electric Company. It provides university, government, and industrial researchers with access to high-performance computing, communications and data-handling for unclassified research. PSC will provide a two day introduction to high performance computing and parallel programming to H<sup>P</sup>C participants. This workshop will introduce participants to the fundamentals of high performance computing, parallel programming, common software packages, and provides practical, hands-on experience in how to write and execute parallel programs. For this audience, the goal is for participants to develop sufficient familiarity with this topic to determine how it might be of value in their research.

The San Diego Supercomputer Center (SDSC)<sup>8</sup> enables scientific discovery and learning through provision of high performance data-intensive computing, analysis, management, and preservation technologies and expertise. Its education programs introduce the next generation users of new computing technologies to the skills and knowledge they will need to explore and discover answers to their generation's challenges. Hallmark programs and products focus on educator professional development (TeacherTECH)<sup>9</sup> and a web portal for integrating data investigation tools and curricular activities into secondary level and college courses. Led by Reagan Moore, Director of the Data-Intensive Computing Environments (DICE)<sup>10</sup> group, and Richard Marciano, Director of the Sustainable Archives and Library, the two-day SDSC *Data Challenges in the Humanities* mini-residence will feature case studies in humanities that highlight data curation and preservation challenges amenable to technology solutions; and community-led technology initiatives that have addressed similar challenges. The SDSC team will introduce an innovative data grid technology that equips users to handle a full range of distributed data management needs, from extracting descriptive metadata, to managing data, to moving it efficiently, sharing data securely with collaborators, publishing it in digital libraries, and archiving data for long-term preservation. One featured technology will be the Integrated Rule-Oriented Data System (IRods), an innovative "rule engine" that lets data collection users more easily accomplish complex data management tasks including validating the trustworthiness of digital repositories and developing community-wide policies to manage data. SDSC will work with the humanities groups to customize IRod and grid technologies for their individual use during the mini-residency and via the virtual community.

Researchers in the social sciences and humanities are increasingly using computers to manage, organize and analyze non-numerical data from textual sources including images, manuscripts, and video. The NCSA two-day mini-residence on Qualitative Data Analytics and Visualization would examine technologies for imaging, image analyses, and environments based on large volumes of data. Computer technologies available to humanities scholars would include (a) integration of historical spatio-temporal data with maps and web-based interfaces (georeferencing, spatial and temporal sampling, sub-setting, tiling and stitching, web-based open layers and server), (b) automated analysis of scans of historical manuscripts (color spaces, image statistics, classification, cropping), (c) 3D imaging using high resolution 2D images of historical artifacts or 2D videos (3D imaging principles, stereopsis, calibration, spectral properties), (d) analyses of large volumes of contemporary PDF documents (PDF document structure, information extraction and cleansing, clustering, versioning) and (e) self-describing executions of analyses using advanced workflow studio (preservation, scripting & workflow, provenance, tagging, distributed data & tools & computers.) Importantly, these collaborative discussions will focus on specific software solutions to the challenges faced by these humanities groups which range from understanding computational requirements using a desktop versus a supercomputer, data presentation formats from desktop visualizations to web-based data browsing and the technological challenges associated with simple image processing applied to large volumes of images including the more complex image analyses executed in real-time. Each humanities research group will creatively analyze, infer, and visualize their data sets, then present their work to the whole group via the virtual community feature of the H<sup>P</sup>C grant.

### **Participants:**

I-CHASS has selected three humanities research groups to be in residence for the grant. Each group will have approximately eight to twelve members from outside the region with an additional seven scholars from local institutions (museums and humanities departments) comprised of senior faculty, researchers, and graduate students in the humanities (Appendix III: List of Participants by Group and Appendix VI: Curriculum Vitae- Humanities Group Leaders).

The Scholarly Community for the Globalization of the "Middle Ages" (SCGMA) Group has been collaborating with SEASR, I-CHASS, the Center for Medieval Studies at the University of Minnesota-

Twin Cities, the Program in Medieval Studies at the University of Texas-Austin, and the Communications Department at the University of California-San Diego since May 2007 to develop a new interdisciplinary scholarly community for globalizing the study of the “Middle Ages”(~500-1500 C.E.)<sup>11</sup> SCGMA has been actively working to create an online infrastructure to support the organization of, and research with, sources in multiple formats and languages available from multiple scholarly disciplines in order to organize large quantities of textual, visual, and aural resources. SCGMA is intended to become a multi-university, multi-nation, disaggregated yet well-coordinated organization spearheading numerous scholarly projects that will challenge the Euro-centrism associated with studying the “Middle Ages.” SCGMA needs to consider adapting existing technologies for its needs and plans to investigate the following: Can a coordinated online resource be developed for students at all levels, from all parts of the globe, already fascinated by aspects of this crucial period but still insufficiently informed about its diversity and range? How can existing databases in many languages and formats, sometimes reflecting different cultural practices, be brought into communication to serve the needs of the developing SCGMA and of students and other interested parties worldwide? H<sup>P</sup>C will allow SCGMA to extend its current use of high performance technologies, resulting from its previous work with ICHASS and SEASR to encompass a more elaborate technological model. The grant offers SCGMA the opportunity to add new technology to its growing infrastructure while simultaneously establishing long-term partnerships.

The University of Southern California's Institute for Multimedia Literacy (IML)<sup>12</sup> has faced a material challenge for the past eight years in realizing one of its primary goals: creating a digital archive system in support of the creation of digital portfolio application.<sup>13</sup> The lack of sufficient computational resources for holding large collections of multimedia resources, most notably its robust digital portfolio of media-rich student projects and faculty teaching resources, has hindered IML's creation of a pedagogical tool for faculty and students. The Humanistic Algorithms project is a collaboration between SEASR, ICHASS, and IML to address this challenge. The project is being imagined in phases, with the first stage to serve as a prototype to be completed by early June. SEASR will use data analytics to extract information from unstructured texts (i.e., raw textual data like websites, etc.) to produce semantic information that can be used to create meta-analyses of scholarly multimedia. From these meta-analysis, Humanistic Algorithms would like to contemplate: What are the components of scholarly multimedia? What is pedagogy in a networked world? How do we collaborate, train faculty, and teach students how to read and compose scholarly multimedia? H<sup>P</sup>C will allow Humanistic Algorithms the opportunity to further their technological efforts by isolating and adapting additional high performance computing technologies that will aid in the development of the digital portfolio application. The series of mini-residencies will allow the group to experiment with new technologies and chart long-term technological planning goals that will enable the program to move forward with its multi-university project.

The HistorySpace Project brings together humanities scholars experimenting with Information Rich Virtual Environments (IRVE) that express combinations of textual, graphic, sonic and three- and four-dimensional forms of expression in order to collaborate on workflows and disciplinary conventions and protocols that will transition humanities scholars from print to virtual media production. With the emergence of XML (Extensible Markup Language), related XML-dialects Like SVG (Scalable Vector Graphics) and X3D, humanities scholars have been able to generate and distribute over the Internet unprecedented datasets and dynamic representations of objects and environments. Combining primary source data with visual imaging and scholarly analysis, IRVEs offer tremendous potential to create elaborate three and four-dimensional renderings of the past. Historians have already begun to appropriate IRVEs as a new method of scholarly communication: the *Aurora Project*, *Virtual Jamestown*,<sup>14</sup> and the *3D Virtual Buildings Project* have begun to experiment with two- and three-dimensional forms of representation to support analysis, expression, and education. Despite these emerging competencies and digital resources, the historical discipline is still not ready to appropriate IRVEs on a widespread basis. When historians generate print-based materials, they rely on established disciplinary criteria to narrate, document, and develop their analysis. Yet, humanities scholars can lay claim to no analogous sets of

workflows and conventions to govern content creation, expression, dissemination and peer review for IRVEs. The HistorySpace Project will conceive, construct and test IRVE workflows, conventions, and accompanying tools to support their operation and expression by humanities scholars. H<sup>P</sup>C will allow the HistorySpace Project, which has been collaborating with the NCSA, to take the next step in its project design. As of June 2008, HistorySpace, in conjunction with participants at NCSA, will have constructed an elaborate network of storyboards and workflow charts that will serve as the iterative, scenario-based design method that will structure the IRVE. This grant will allow HistorySpace to consult with high performance computing specialists in order to refine their IRVE methodological structure, consider the integration and adaptation of additional high performance computing tools, and begin construction of its first prototype IRVE.

### **Impact and Evaluation:**

Findings from this project will be disseminated in traditional and innovative ways. We will encourage and maintain collaboration among humanities and high performance computing participants by creating a web-portal that allows electronic dissemination and maintains a constant web-based presence. The primary goal of disseminating experience, breakthroughs and learned lessons will be achieved by putting the mini-residences online via webcast. The secondary goal will be facilitated by the inclusion of modules embedded in the portal that allows activities tracking (provenance data), collaboration support (including blogs, chat, and wikis), and networking support. Creating a multi-way networked activity centered on the digital humanities, the experience, breakthroughs, and lessons learned from each project will be disseminated to the wider humanities academic audience and the general public through our virtual community in addition to the more traditional online papers, journal articles and research reports. Our ambition is to achieve continual and energetic discussion and collaboration for each group and as a collaborative. H<sup>P</sup>C involves a unique combination of assets: 1) it will further the research efforts of the selected humanities research groups; 2) it will allow for the refinement of technologies by computational scholars and humanities users; 3) it will create a collaborative infrastructure and virtual community accessible to scholars across humanities. The mini-residences and the two-day conference, which will be webcast, will culminate in an ever-evolving planning memorandum, "Coordinating High Performance Computing and the Digital Humanities." It will outline long-range technological planning goals that can be undertaken by the partnership and each humanities group upon completion of the grant and aid participants in charting the systematic planning and developmental goals needed to integrate these new technologies within their project. Beyond group participants, these long-range planning memoranda will continually evolve as new technologies, methods, and humanities groups join the collaboration and its parallel virtual community. In serving not just invited participants but also scholars interested in the humanities and digital technologies, H<sup>P</sup>C captures the underlying intent of the National Endowment for the Humanities grant programs: to interest people in the Humanities and aid them in their quest to more fully understand human life and experiences.

Within the confines of H<sup>P</sup>C, our evaluation process will be conducted via evaluation tools provided by the University of Illinois' Center for Instructional Research and Curriculum Evaluation (CIRCE) and the internal evaluation staff at NCSA. CIRCE has evaluated programs at all educational levels and will bring together humanities scholars experienced in evaluation to aid in the assessment of H<sup>P</sup>C. They will collect data via surveys, content assessment instruments, and interviews over the course of the grant. Participants will be asked to assess the curriculum, a series evaluation to recommend strategies to improve the virtual

workshop and community, and a final qualitative evaluation that will track the continued use of these technologies and the efforts to secure additional grant funding dollars. Program administrators at each center will also be surveyed to consider the ways in which the collaboration can be strengthened and changed over time to better serve the technological participants and their needs.

### **Staff, Faculty, and Consultants (Appendix V: Curriculum Vitae- I-CHASS Faculty and Staff):**

*Principal Investigator:* Orville Vernon Burton is Director of the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois, where he is Professor of History and African American Studies and serves as a Senior Research Scientist at the National Center for Supercomputing Applications, where he is Associate Director for Humanities and Social Sciences. Burton is the author of more than a hundred articles and the author or editor of fourteen books (one of which is on cd-rom), including *In My Father's House Are Many Mansions: Family and Community in Edgefield, South Carolina* and *The Age of Lincoln*.

*Co-Principal Investigator:* Kevin Franklin is Executive Director of the Institute for Computing in Humanities, Arts and Social Sciences (I-CHASS) and Senior Research Scientist for the National Center for Supercomputing Applications (NCSA). He is the former Executive Director of the University of California Humanities Research Institute and Deputy Director of the University of California San Diego Supercomputer Center. Dr. Franklin serves as co-chair for the Humanities, Arts and Social Science Research Group for the Open Grid Forum and on the Advisory Board for the Worldwide University Network Grid Advisory Committee. He designed and implemented the University of California systemwide online research proposal application tool and the Humanities, Arts and Social Science Grid (HASSgrid). He is a co-founder of the Humanities, Arts, Science and Technology Advanced Collaboratory (HASTAC).

*Project Manager:* Simon Appleford received a Masters of Arts in Modern American History and a Masters of Literature from the University of St. Andrews, Scotland before joining NCSA in 2005. His interests in digital technologies and American history have led to several publications including articles in *CTWatch Quarterly* and *Toward the Meeting of the Waters: Currents in the Civil Rights Movement in South Carolina* (University of South Carolina Press, 2007.) Simon was the principal organizer of e-Science for Arts and Humanities Research: Early Adopters Forum (2007), *Spatial Thinking in the Social Sciences and Humanities* (2006), and *Computing in Humanities, Arts, and Social Science* (2005). He is currently completing his Doctorate of Philosophy in History at the University of Illinois while serving as Project Manager at I-CHASS. Simon Appleford will be responsible for project coordination and collaboration between the high performance computing centers and the humanities groups including the maintenance of the virtual community and the grant's findings.

*Senior Project Scientist:* Peter Bajcsy received his Master of Science in Electrical Engineering from the University of Pennsylvania and his Doctorate in Electrical and Computer Engineering from the University of Illinois. Peter and the SEASR group have been investigating and developing solutions to real life problems in the application areas of remote and airborne sensing, geo-spatial information systems (GIS), target and scene modeling from multi-spectral and synthetic aperture radar (SAR) imagery, bio-informatics and health informatics, microscopy and medical image processing, automated information extraction and organization from large size image scans and PDF documents, 3D imaging and advanced sensor environments. He is currently employed in multiple positions at the University of Illinois: as the Associate Director for Data Analytics and Pattern Recognition at I-CHASS, as Adjunct Assistant Professor in the Electrical and Computer Engineering and Computer Sciences Departments, and as a Research Scientist in Image Spatial Data Analysis (ISDA) at NCSA. Peter Bajcsy will design and implement the education and training activities that will be hosted at NCSA.

*Project Scientist:* Alex Yahja earned his Ph.D. degree in computation, organizations and society from Carnegie Mellon University, Pittsburgh, PA, in 2006, and currently works on the interface between technologies and humanities, arts and social sciences at the National Center for Supercomputing Applications in Urbana, IL. The problems he has worked on include social drivers in disaster response, collaboration across disciplines, network-based recommendation, mapping of research activities, and semantics-based specification and collaboration. He received two M.Sc. degrees, one in engineering and public policy and one in robotics from Carnegie Mellon University, Pittsburgh, PA, in 2004 and 2000 respectively.

*Project Scientist:* Alan Craig has focused his career on the interface between humans and machines. He has been involved in many different capacities related to scientific visualization, virtual reality, data mining, multi-modal representation of information, and collaborative systems during his career at the National Center for Supercomputing Applications where he has worked for the past twenty years. Dr. Craig is co-author of the book *“Understanding Virtual Reality”*, published by Morgan Kaufmann Publishing, and author of the forthcoming book, *“Using Virtual Reality”*.

*Project Coordinator:* Jim Onderdonk is Associate Director for Education and Outreach with the Institute for Computing in Humanities, Arts and Social Sciences (I-CHASS). He also serves as Head, Conferences and Institutes (C&I), one of the divisions in the Office of Continuing Education at the University of Illinois at Urbana-Champaign. His undergraduate degree is from the College of William and Mary and his masters and doctorate are from Old Dominion University.

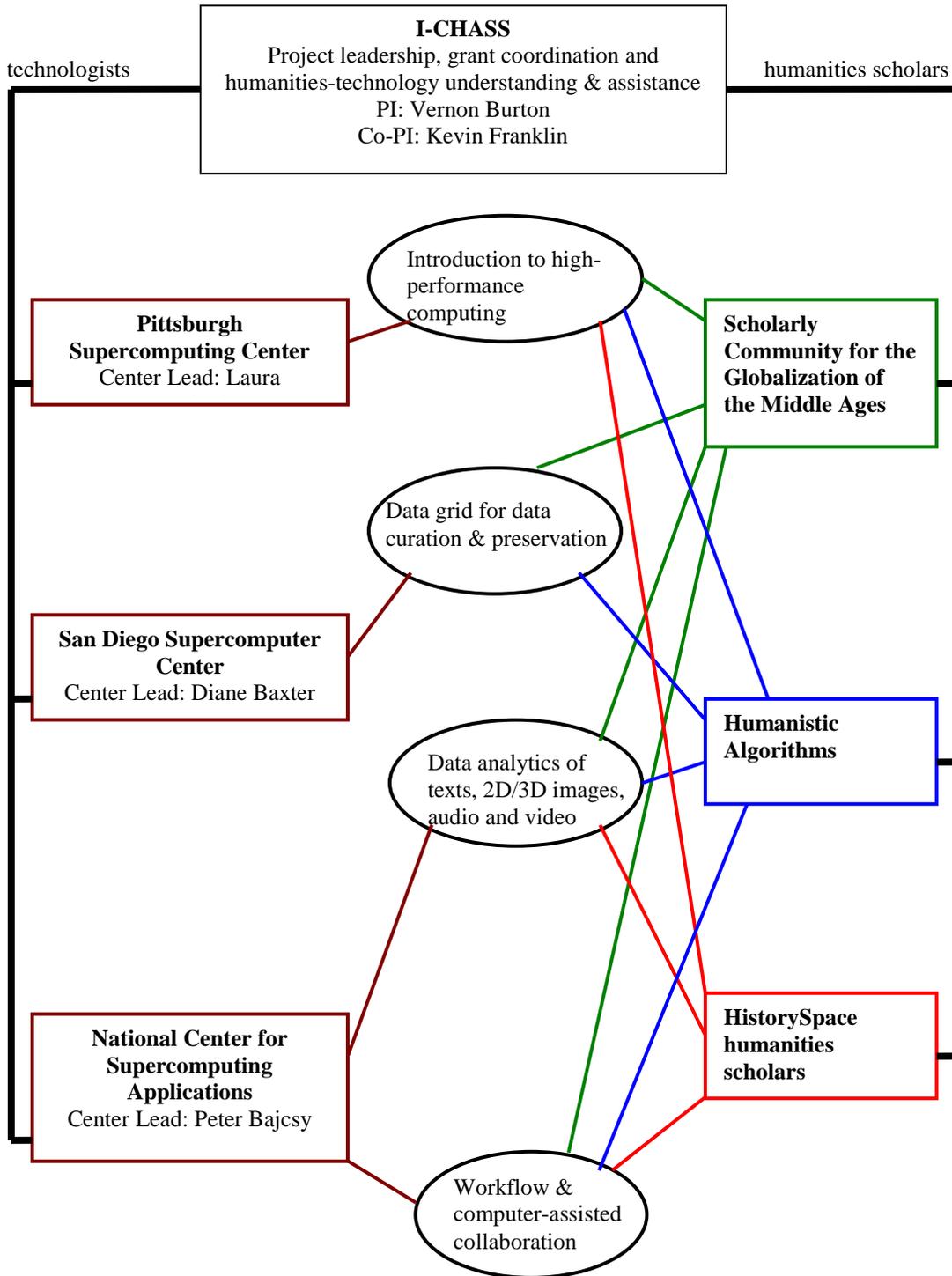
## Appendix I: References

- 1.) The NEH has recognized the challenge associated with digital technologies and high performance computing for the humanities. In July of 2007, the NEH sponsored the “Humanities High Performance Computing” Conference which included inviting participants from High Performance Computing Centers and Digital Humanities Centers including the Institute for Computing in the Humanities, Arts, and Social Science along with representative from the San Diego Super Computing Center, The Institute for Advanced Technology and the Humanities at the University of Virginia, the Department of Energy and the National Science Foundation. That meeting served to open dialogues between these groups about the possibilities of a multi-center multi-group collaboration.
- 2.) We use the term “Humanities High Performance Computing” intentionally to signal an investment (structural, computational, and resource-based) in the extension of I-CHASS to serve as a portal for humanities scholars to receive technical support, access to high performance computing, and products and services associated with the digital technologies. We envision that additional centers will join this grant in later stages and I-CHASS will serve as a national entry point for humanities, arts, and social science researchers to receive high performance computational services at all levels of expertise from beginner to the most advanced humanities computing projects.
- 3.) The call for multi-center collaborations has been offered in the sciences as well in the humanities. “No one agency can – or should – carry all the weight of ensuring that our scientists have the computational tools they need to do their job”, spoke Dr. Raymond L. Orbach, Director of Science at the Department of Energy during his 2003 testimony before the House Committee on Science, “yet duplication of effort must be avoided.” See [http://www.er.doe.gov/sub/speeches/Congressional\\_Testim/7\\_16\\_03\\_testimony.htm](http://www.er.doe.gov/sub/speeches/Congressional_Testim/7_16_03_testimony.htm) for his full comments.
- 4.) <http://valley.vcdh.virginia.edu>
- 5.) The Great Lakes Consortium for Petascale Computation is a collaboration among colleges, universities, national research laboratories, and other educational institutions that facilitates the widespread and effective use of petascale computing by developing new computing software, applications, and technologies. A “petascale” system is expected to be able to make arithmetic calculations at a sustained rate in excess of a sizzling 1,000-trillion operations per second (a "petaflop" per second) to help investigators solve some of the world's most challenging research problems. Please consult: <http://www.greatlakesconsortium.org/> for more information.
- 6.) I3 was established in 2007 at the University of Illinois to foster multi-disciplinary collaboration, support joint academic appointments, offer courses and academic programs, and sponsor research and technology development in informatics and its applications.
- 7.) <http://www.psc.edu>
- 8.) <http://www.sdsc.edu/>
- 9.) <http://education.sdsc.edu/teachertech/>
- 10.) <http://education.sdsc.edu/discoverdata/>
- 11.) For examples of the types of scholarship the SCGMA group is relying upon please consult: Thomas T. Allsen, *Culture and Conquest in Mongol Eurasia* (Cambridge, Eng.: Cambridge University Press, 2001); Geraldine Heng, "An Experiment in Collaborative Humanities: Imagining the World, 500-1500 C.E.," *Association of Departments of Foreign Languages (ADFL) Bulletin* 38 (3) and 39 (1), Spring-Fall 2007, 20-28; Peter R. Schmidt, and T. Patterson, *Making Alternative Histories: The Practice of Archaeology and History in Non-Western Settings* (Santa Fe: School of American Research, 1995); Susan Whitfield, and Ursula Sims-Williams, *The Silk Road: Trade, Travel, War, and Faith* (Chicago: Serindia Publications, 2004).
- 12.) <http://www.iml.usc.edu/>
- 13.) For information of digital portfolio applications and their relationship to humanities learning please

consult: The Electronic Portfolio Consortium, association of individuals from 50 countries and almost 600 higher education and IT commercial institutions, who are interested in the development of academic ePortfolio software systems and the establishment of standards for such systems, <http://www.eportconsortium.org/>; The Open Source Portfolio Initiative, a community of individuals and organizations collaborating on the development of the leading non-proprietary, open source electronic portfolio software, <http://www.osportfolio.org/>; Terence Love and Trudi Cooper, "Designing Online Information Systems for Portfolio-Based Assessment: Design Criteria and Heuristics," InSITE, Informing Science, June 2003.

14.) <http://www.virtualjamestown.org/>

**Appendix II: Collaboration Diagram**



Note: **brown** lines denote provision and **red/green/blue** lines denote priority interests

### **Appendix III: List of Participants By Group**

#### **High Performance Computing Centers: National Center for Supercomputing Applications**

Appleford, Simon	Project Manager, Institute for Computing in the Humanities, Arts, and Social Science; Graduate Student, Department of History, University of Illinois
Bajcsy, Peter	Associate Director for Data Analytics and Pattern Recognition, Institute for Computing in the Humanities, Arts, and Social Science.
Burton, Vernon	Director of the Institute for Computing in the Humanities, Arts, and Social Science, Professor of History, African American Studies, and Sociology; Senior Research Scientist and Associate Directory for Humanities and Social Sciences at the National Center for Supercomputing Applications (NCSA).
Craig, Alan	Associate Director for Human-Computer Interaction, Institute for Computing in the Humanities, Arts, and Social Science.
Franklin, Kevin	Executive Director of the Institute for Computing in the Humanities, Arts and Social Sciences; Senior Research Scientist at the National Center for Supercomputing Applications (NCSA).
Guiliano, Jennifer	Ph.D. Candidate, Department of History, University of Illinois; Graduate Assistant, Institute for Computing in the Humanities, Arts, and Social Science; Graduate Assistant, Department of History.
Onderdonk, James	Associate Director for Education and Outreach, Institute for Computing in the Humanities, Arts, and Social Science.
Seawell, Stephanie	Graduate Student, Department of History, University of Illinois; Graduate Assistant, Institute for Computing in the Humanities, Arts, and Social Science; Graduate Assistant, Department of History.
Yahja, Alex	Assistant Director in Modeling, Institute for Computing in the Humanities, Arts, and Social Science.

#### **Pittsburgh Supercomputing Center:**

McGinnis, Laura F.	Project Manager, Data and Information Resource Services, Pittsburgh Supercomputing Center
Brown, Shawn T.	Sr. Scientific Specialist, Scientific Applications and User Services, Pittsburgh Supercomputing Center
Maiden, Tom	User Support and Outreach Specialist, Scientific Applications and User Services, Pittsburgh Supercomputing Center
Urbanic, John	Staff Computational Science Consultant, Strategic Applications, Pittsburgh Supercomputing Center

#### **San Diego Supercomputer Center:**

Baxter, Diane	Director of Education, San Diego Supercomputer Center
Moore, Reagan	Director, Data-Intensive Computing Environments (DICE) group, San Diego Supercomputer Center
Marciano, Richard	Director, Sustainable Archives and Library, San Diego Supercomputer Center

## Humanities Groups:

### The Scholarly Community for the Globalization of the “Middle Ages” (SCGMA) Group:

Allsen, Thomas	Professor Emeritus, College of New Jersey
Asher, Catherine B.	Associate Professor of Art History, University of Minnesota
Aytes, Ayhan	Visual Media Researcher and Graduate Student in Communications, University of California- San Diego
Boone, Jim	Associate Professor of Anthropology, University of New Mexico
Goldberg, David Theo	Director, University of California Humanities Research Institute (UCHRI); Professor of Comparative Literature and Criminology, Law, and Society at the University of California, Irvine.
Hart, Roger	Assistant Professor of History, University of Texas at Austin.
Heng, Geraldine	Director of Medieval Studies, Associate Professor of English; Holder of the Perceval Endowment in Medieval Romance, Historiography, and Culture at the University of Texas at Austin.
Ilnitchi, Gabriela	Assistant Professor of Musicology, University of Minnesota.
Kea, Ray A.	Professor of History, University of California, Riverside
Klieman, Kairn A.	Associate Professor of History, University of Houston, Texas.
Larkin, Margaret.	Professor of Arabic Literature, Department of Near Eastern Studies, University of California, Berkeley.
Liu, Xinru	Assistant Professor of Early Indian History and World History, College of New Jersey.
Martin, Will	Graduate Student, Department of Information Science, University of Texas, Austin.
McIntosh, Susan K.	Professor of Archaeology, Rice University.
Noakes, Susan	Director, Center for Medieval Studies, Professor of French and Italian, University of Minnesota, Twin Cities.
Phillips Jr., William D.	Director, Center for Early Modern History, Professor of History, University of Minnesota
Ragnow, Marguerite	Curator, James Ford Bell Library, University of Minnesota.
Saldanha, Arun	Assistant Professor, Geography, University of Minnesota.
Schmidt, Peter R.	Professor of Anthropology, University of Florida
Spellberg, Denise	Associate Professor of History and Middle Eastern Studies, Associate Director of Medieval Studies, University of Texas.
Urbam, Tomislav	Data and Information Service Group, Texas Advanced Computing Center.
Waltner, Ann B.	Director, Institute for Advanced Study, University of Minnesota.
Widner, Mike	Graduate Student, Medieval Literature, University of Texas, Austin.
Wilcox, Rebecca	Graduate Student

### The Humanistic Algorithms: Semantic Networks in Multimedia Scholarship Group:

Arroyo, Sarah J.	Assistant Professor of English, California State University at Long Beach
Ball, Cheryl	Assistant Professor, English, Illinois State University; Editor of Kairos, refereed online journal exploring the intersections of rhetoric, technology, and pedagogy

Blalock, Glenn	Assistant Professor of Composition and Rhetoric, Department of English, Baylor University; Creator of CompPile, a scholarly database of journals in rhetoric and compositions.
Callahan, Vicki	Founder of the Conceptual Studies in Media Arts Production at the Peck School of the Arts (Film Dept), University of Wisconsin, Milwaukee
Franklin, Kevin	Executive Director, Institute for Computing in the Humanities, Art, and Social Science, University of Illinois
Haswell, Rich	Professor Emeritus, Department of English, Texas A & M, Corpus Christi; Creator of CompPile, a scholarly database of journals in rhetoric and compositions.
Hawisher, Gail	Professor of English and Director of the Center for Writing Studies, University of Illinois.
Holmes-Wong, Deborah Kuhn, Virginia	Archivist for Digital Media, University of Southern California Associate Director, Institute for Multimedia Literacy, University of Southern California
Selfe, Cynthia L.	Humanities Distinguished Professor, Department of English, Ohio State University
Vitanza, Victor J.	Professor of English and Director of the Ph.D. Program in Rhetorics, Communication, and Information Design, Clemson University
Wright, Elijah	Adjunct Lecturer and Graduate Student, School of Library and Information Sciences, Indiana University
Wysocki, Anne F.	Associate Professor English, University of Wisconsin, Milwaukee

**The HistorySpace Project: Information Rich Virtual Environments for Historical Scholarship Group:**

Ayers, Edward	President, University of Richmond; Aurora Project; <i>Valley of the Shadow</i> Project.
Bonnett, John	Assistant Professor, Humanities Institute, Department of History, Brock University (Canada)
Denard, Hugh	Professor, Centre for Computing in the Humanities, Director, King's Visualization, Lab King's College London; Editor, <i>Didaskalia</i> ; Member, <i>The Pompey Project</i>
Dunae, Patrick	Professor, Department of History, University of Victoria; Co-Director, Vi.History.ca (online); editor of The Homeroom, British Columbia's online history of education website.
Dunn, Stuart	Research Associate, Centre for Computing in the Humanities, King's College London.
Gilliland, Jason	Assistant Professor, Department of Geography; Director, Urban Development Program, University of Western Ontario.
Lutz, John	Professor, Department of History, University of Victoria; Co-Director, Great Unsolved Mysteries in Canadian History (online); Co-Director, Victoria's Victoria (online); Co-Director, Vi.History.ca (online); Co-Director, Who Killed William Robinson? Race Justice and Settling the Land (online).
Thomas III, William G.	John and Catherine Angle Professor in the Humanities Department of History, University of Nebraska; Aurora Project

**Appendix IV: Curriculum Workplan Itinerary**

Mini-Residence

<b><i>Mini-Residences</i></b>		
<b>Day 1</b>		
<b>Tues.</b>	<b>Welcome to the Workshop</b>	
7:00 PM	Welcome Dinner	Primary Collaborator- HP Center and HP Center Administration
<b>Day 2</b>		
<b>Wed</b>	<b>Technology and the Humanities</b>	
9:30-9:45 AM	Introduction to the Workshop Series	Primary Collaborator- HP Center
9:45-10:30 AM	Mini-Residence Research Goals and Objectives	Primary Collaborator- HP Center
10:30-11:00 AM	Introduction to the Research Project, their Goals, and Objectives	Humanities Research Group
11:00-12:15 AM	Lunch	
12:30-3:00 PM	Technology and your Research Group	HP Center technology staff
3:00-3:15 PM	Break	
3:15-4:45 PM	Technology and your Research Group pt. 2	HP Center technology staff
5:30 PM	Dinner	
<b>Day 3</b>		
<b>Thurs</b>	<b>Adapting High Performance Computing Technology</b>	
8:30-10:00	Adapting High Performance Computing Technology	HP Center technology staff
10:00-10:15	Break	
10:15-12:00	Adapting High Performance Computing Technology pt. 2	HP Center technology staff
12:00-1:00	Lunch	
1:00-3:30	Adapting High Performance Computing Technology pt. 3	HP Center technology staff
3:30-3:45	Break	
3:45-5:00	Long-Term Technology Planning	Primary Collaborator- HP Center

Summer Conference

<b><i>Conference</i></b>		
<b>Day 1</b>	<b>Humanities Research</b>	
8:30-8:45	Welcome Remarks	Vernon Burton, I-CHASS
8:45-10 AM	Humanities Research Group 1 Presentation	<i>The Scholarly Community for the Globalization of the "Middle Ages" Group, Susan Noakes, lead.</i>
10-11:15 AM	Humanities Research Group 2 Presentation	<i>The Humanistic Algorithms: Semantic Networks in Multimedia Scholarship Group, Virginia Kuhn, lead.</i>
11:30- 1 PM	Lunch	
1- 2:15 PM	Humanities Research Group 3 Presentation	<i>The Human- and Cyber-infrastructures to Study and Combat Violence Group, Heidi Beirich, lead.</i>
2:15-3:30 PM	Humanities Research Group 4 Presentation	<i>The HistorySpace Project: Information Rich Virtual Environments for Historical Scholarship, John Bonnett, lead.</i>
3:30-3:45 PM	Break	
3:45-4:30	Discussion: and "Issues of Technology and the Digital Humanities"	All participants
6:30-7:30 PM	Dinner	
<b>Day 2</b>	<b>The Digital Future of Humanities Research</b>	
9-11:30	"Coordinating High Performance Computing and the Digital Humanities"	All participants
11:30- 1 PM	Lunch	
1-3 PM	Open Discussion	All participants
3- 4 PM	Evaluation	CIRCE Staff
4:00-4:15 PM	Closing Remarks	Kevin Franklin, I-CHASS

**Appendix V: Curriculum Vitae**  
**I-CHASS Faculty, Staff, and Consultants**

**ORVILLE VERNON BURTON**

**Professional Preparation**

Furman University History B.A. 1969  
Princeton University History M.A. 1971  
Princeton University History Ph.D. 1976

**Appointments**

2004 – Present Director, Institute for Computing in Humanities, Arts, and Social Science  
2003 – Present Associate Director, Humanities and Social Sciences, National Center for Supercomputing Applications (NCSA) at University of Illinois Urbana Champaign (UIUC)  
2001 – Present Executive Director, Lowcountry and Atlantic World Program, College of Charleston  
2000 – 2001 Mark W. Clark Visiting Distinguished Chair, The Citadel  
1995 – Present Senior Research Scientist, NCSA  
1993 – 2003 Head Initiative for Social Science and Humanities at NCSA  
1991 – 1995 Professor, NCSA  
1989 – Present Professor, History, UIUC  
1989 – Present Professor, Sociology, UIUC  
1988 – 1991 Adjunct Professor, NCSA  
1987 – Present Faculty Affiliate, Afro-American Studies and Research Program, UIUC  
1986 – 1988 Faculty Affiliate, NCSA  
1986 -- Present Professor Campus Honors Program, UIUC  
1982 – 1989 Associate Professor, History, UIUC  
1981 – Present Graduate College Statistics Faculty, UIUC  
1976 – 1982 Assistant Professor, History, UIUC  
1974 – 1976 Instructor, History, UIUC  
1972 – 1974 Assistant Master, Woodrow Wilson Residential College, Princeton University  
1971 – 1972 Instructor, Mercer County Community College, New Jersey

**Publications**

editor, *Computing in the Social Sciences and Humanities*. Urbana: University of Illinois Press, 2002 (Recognized as a Choice” Outstanding Academic Book, 2003”).  
(edited with David Herr and Terence Finnegan) *Wayfarer: Charting Advances in Social Science and Humanities Computing*. Urbana: University of Illinois Press, 2002. CD-ROM contains more than 65 essays and research and teaching applications, including illustrative interactive multimedia materials.  
(with Ian Binnington, David Herr, and Matthew Cheney) “Computer Mediated Learning Environments: How Useful Are They?,” *AHR Perspectives: Newsmagazine of the American Historical Association* 41:1 (January 2003): 14, 22 (More detailed Carnegie Report as “Historians Face the E-Future: Findings from the Carnegie Scholar Survey on Computer Mediated Learning Environments,” at AHA Website [www.theaha.org/perspectives/issues/2003/0301/0301not3.cfm](http://www.theaha.org/perspectives/issues/2003/0301/0301not3.cfm)) and expanded version published as “What Difference Do Computers Make? History, Historians, and Computer-Mediated Learning Environments,” *History Computer Review* 19 (Spring 2003): 98-103.  
(with Simon Appleford) “The Illinois Center for Computing in Humanities, Arts, and Social Science,” *Cyberinfrastructure Technology Watch Quarterly (CTWatch)* <http://www.ctwarch.org>. May, 2007.  
"Complementary Processing: A Supercomputer/Personal Computer U.S. Census Database Project" in

Supercomputing 88, vol. 2 Science and Applications. Edited by Joanne L. Martin and Stephen Lundstrom. Washington, D.C.: IEEE Computer Society Press, 1990, pp. 167-177.

*The Age of Lincoln*. New York: Hill and Wang, 2007. Recipient of the 2007 Heartland Award.

*In My Father's House Are Many Mansions: Family and Community in Edgefield, South Carolina*.

Chapel Hill: University of North Carolina Press, 1985 (paperback edition 1987; 5th printing 1998). *In My Father's House* was nominated by the University of North Carolina Press for the Pulitzer Prize. Two professional academic Associations have featured this book in sessions at their annual meetings: Social Science History Association, 1986; Southern Historical Association, 1987.

"American Digital History," *Social Science Computer Review* 23: 2 (Summer 2005): 206-220.

(with Terence Finnegan, Peyton McCrary, and James W. Loewen) "South Carolina" chap. 7, pp. 191-232, 420-432 in *The Quiet Revolution in the South: The Impact of the Voting Rights, 1965-1990*. Edited by Chandler Davidson and Bernard Grofman. Princeton: Princeton University Press, 1994. (Winner of the 1995 Richard F. Fenno Prize, Legislative Studies Section, American Political Science Association)

"Reaping What We Sow: Community and Rural History," Presidential address in *Agricultural History* 76: 4 (Fall 2002): 631-58.

#### synergistic activities

U.S. Professor of the Year, Outstanding Research and Doctoral Universities Professor (Council for Advancement and Support of Education and Carnegie Foundation for the Advancement of Teaching) 1999 recognized for the introduction of Information Technology into the Classroom and the study of Diversity in the Classroom; American Historical Association Eugene Asher Distinguished Teaching Award, 2003; appointed Organization of American Historians Distinguished Lecturer, 2004-07

Certificate of Excellence from the Carnegie Academy for the Scholarship of Teaching and Learning for Work that Advances the Practice and Profession of Teaching In Support of Significant Student Learning, June 28, 2001.

Named one of the first three University of Illinois at Urbana Champaign University "Distinguished Teacher/Scholar" 1999.

Organized number of meetings. (selected) 1993: organized, hosted, and chaired the annual meeting of the Conference on Computing for the Social Sciences at the National Center for Supercomputing Applications at Illinois Edited special issue of *Social Science Computer Review* 12:2 (Summer 1994) from papers presented at conference. 2003: organized three meetings-A Workshop on Diversity and Racism in the Classroom for university faculty, public school teachers, and the community (January as co-chair of Martin Luther King, Jr. Week at UIUC); organized a Conference on the Scholarship of Teaching and Learning for Illinois Faculty (Jan.); organized a conference on the Civil Rights Movement in South Carolina at The Citadel in Charleston, S.C. (March) As Director of Institute for Computing in Humanities, Arts, and Social Science regularly host workshops on cyber infrastructure and information technology- 2007 hosted workshops on GIS with Luc Anselin and Grid Computing with English e-science scholars. Currently preparing books and reports from each of these workshops.

#### collaborators and other affiliations

1. Collaborators: Simon Appleford, UIUC; Ian Binnington, Allegheny College., Richard Braatz (UIUC), Beatrice Burton, University of Georgia; Georganne B. Burton, spouse; Matthew Cheney (UIUC); Terence Finnegan, William Paterson U.; David Herr, St. Andrews College; Eric Jakobsson (UIUC), Mark Kornbluh (Michigan State U.), Winfred Moore, The Citadel; David O'Brien, UIUC; James Onderdonk, UIUC; Richard Pate (Danville Community College); Deanna Raineri (UIUC),

2. Advisors: F. Sheldon Hackney, U. of Pennsylvania and James M. McPherson, Princeton U.

## KEVIN FRANKLIN

Executive Director, Institute for Computing in the Humanities, Arts and Social Sciences  
Senior Research Scientist, National Center for Supercomputing Applications  
University of Illinois at Urbana Champaign  
1205 W. Clark Street  
Urbana, IL 61801  
Phone: 217-265-4044  
Cell: 858-336-8285

### Professional Preparation

Old Dominion University Psychology B.S 1982  
Old Dominion University Education M.S. 1984  
University of San Francisco Organization and Leadership Ed.D. 1993

### Appointments

Executive Director, Institute for Computing in the Humanities, Arts and Social Sciences, University of Illinois, Senior Research Scientist National Center for Supercomputing Applications, Urbana Champaign, July 2007 – present  
Executive Director, University of California Humanities Research Institute, UC Irvine, 2002-2007  
Deputy Director, University of California Supercomputer Center, UC San Diego 2000-2001  
Executive Director, Nonprofit Ventures, Inc, 1998-2000  
Interim Executive Director, Summerbridge National, 1997-1998  
Senior Fellow, San Francisco State University, Urban Institute, 1993-1997  
Director, San Francisco State University, Urban Scholars Program, 1992-1994  
Founding Executive Director, Multicultural Alliance, 1989-1997

### Publications

Co-Editor, Cyberinfrastructure Technology Watch Quarterly Journal, May 2007  
HASS Editor, GridToday, News and Information for Global Grid Communities, 2003-  
HASS Editor, HPCWire, News and Information for High Performance Computing Communities, 2003  
Authored article, “Developing Teachers of Color”, Independent School Magazine, Winter 1991  
Featured in interviews or profiles in New York Times, Los Angeles Times, United Airlines Hemispheres Magazine, Boston Globe, Connecticut Journal, Milwaukee Journal Sentinel, Marin Independent Journal, Essence Magazine, Black Issues in Higher Education, National Public Radio, Cable News Network, KPIX Bay Area Channel 5

### Synergistic Activities

Conference Program Committees: Semantic and Knowledge Grid Conference, China 2006. National Center for eSocial Science Annual Conference, United Kingdom 2005, Latin American VIP Grid Summit, Costa Rica 2005, Richard Tapia, Diversity in Computing Conference 2001, U.S. Department of Education, Teacher Diversity Conference 1995.  
Community-based Committees: California State Advisory Committee to the United States Commission on Civil Rights 1995-2004; Strategic Advisor, Costa Rica-United States Foundation, Costa Rica 2006- present; Director, Cyberinfrastructure Summer Institute for Humanities, Arts and Social Sciences, UCSD, 2006-2007; Chairperson, Executive Committee, Latin American Grid Alliance, CeNAT, Costa Rica 2003-present; Co-Chairperson, Humanities, Arts and Social Science Research Group, Global Grid Forum, 2003-present; Vice-Chair, UC Office of the President, Humanities, Arts, Social Sciences Technology Council, 2004-2007; Worldwide University

Network Grid Advisory Committee, United Kingdom, 2003-present; Trustee, Branson School, Ross CA 1995-1997; Trustee, Marin Horizon School, Marin CA 1992-1996; Trustee, Phillips Brooks School, Palo Alto, CA 1991-1997; Trustee, East Bay French American School, Oakland 1993-1995; Advisory Board Member, School of Education, San Francisco State University 1995-1997; Advisory Board Member, School of Education, University of San Francisco 1996-1997; Chairperson, Teacher Certification Committee, San Francisco State University 1993-1994; Fellowship Advisory Committee, Rockefeller Brothers Fund 1995.

Technical Leadership: Chief-Architect, “FastApps” UC System-wide Online Research Proposal Applications System 2002; Co-Lead, Testbed for Redlining Archives of California’s Exclusionary Spaces UC Irvine, 2006-2007.

Founder/Director, Community Collaborations: Director, Humanities, Arts and Social Science Grid, UC Irvine, 2004-2007; Co-Founder, Humanities, Arts, Science and Technology Advanced Collaboratory, UC Irvine, 2002.

Lead Fundraiser: Fundraiser for successful grants for totaling \$53 million from industry, agencies and foundations including Mobil, ARCO, Kellogg Foundation, Philip Morris Companies, Kraft Foods, Levi Strauss, Hewlett Packard, ORCALE, Mary A. Crocker Trust, Danforth Foundation, Gap Foundation, Hearst Foundation, Rockefeller Brothers Fund, Milwaukee Public Schools, Independent Schools Associations, Corporation for National Service, Los Angeles Jewish Community Foundation, Institute of Museum and Library Sciences.

### **SIMON J. APPLEFORD**

Project Manager, Institute for Computing in Humanities, Arts, and Social Science (ICHASS)  
National Center for Supercomputing Applications  
University of Illinois at Urbana Champaign  
1205 W. Clark Street  
Urbana, IL 61801  
Phone: 217-265-4044  
Cell: 858-336-8285

### **Professional Preparation**

University of St Andrews Modern History M.A. 2000  
University of St Andrews Modern American History M.Litt. 2001  
University of Illinois at Urbana-Champaign, Ph.D. (in progress) in History, 2007-present

### **Appointments**

Project Manager, University of Illinois Institute for Computing in the Humanities, Arts and Social Sciences, Urbana Champaign, February 2007 – present  
Visiting Project Specialist, National Center for Supercomputing Applications, 2005-2007  
Production Editor, Omegatype Typography, 2004-2005  
Senior Technology Specialist, FedEx Kinko’s, 2003-2004  
Publishing Assistant, Leckie & Leckie, 2001-2002

### **Publications**

(with Burton, V. and Onderdonk, J.) “A Question of Centers: One Approach to Establishing a Cyberinfrastructure for the Humanities, Arts, and Social Sciences,” *CTWatch Quarterly*, Volume 3, Number 2, May 2007.  
(with Burton, B. and Burton, V.), “Seeds in Unlikely Soil: The Briggs v. Elliott School

Desegregation Case” in *Toward ‘the Meeting of the Waters’: Currents in the Civil Rights Movement in South Carolina*, ed. Vernon Burton and Winifred B. Moore, Jr., University of South Carolina Press, 2007.

### **Synergistic Activities**

Project Manager for RiverWeb Project ([www.riverweb.uiuc.edu](http://www.riverweb.uiuc.edu))  
Organized several conferences and workshops for I-CHASS: e-Science for Arts and Humanities Research: Early Adopters Forum, 2007; Spatial Thinking in the Social Sciences and Humanities, 2006; Computing in Humanities, Arts, and Social Science, 2005.

### **Collaborators**

(a) Collaborators: Vernon Burton, UIUC; Beatrice Burton, University of Georgia; Kevin Franklin, UIUC; Mark Kornbluh (Michigan State University); Kalev Leetaru (UIUC); David O’Brien, UIUC; James Onderdonk, UIUC; Troy Smith, UIUC

(b) Advisors: Vernon Burton, UIUC; Stephen Spackman, University of St. Andrews; Timothy Minchin, La Trobe University

### **PETER BAJCSY**

#### Professional Preparation.

Slovak Technical University, Bratislava, Czechoslovakia  
Major: Technical Cybernetics and Measurement Techniques  
Degree & Year: Diploma Engineer, 1987  
University of Pennsylvania, Philadelphia, PA  
Major: Electrical Engineering  
Degree & Year: Master of Science, 1994  
University of Illinois, Urbana-Champaign, IL  
Major: Electrical and Computer Engineering  
Degree & Year: Doctor of Philosophy, 1997

#### Appointments.

2003 – Present Adjunct Assistant Professor ECE, UIUC  
2002 – Present Adjunct Assistant Professor CS, UIUC  
2001 – Present Research Scientist NCSA  
1998 – 2001 Senior Scientist SAIC/DEMACO, Inc., Champaign, IL.  
1997 - 1998, Senior Software Engineer Cognex Corporation, Acumen Products Group, Portland, OR

#### Publications. (Most Recent)

1. R. Kooper, A. Shirk, S.-C. Lee, A.Y. Lin, R. Folberg and P. Bajcsy, "3D medical volume reconstruction using Web services." *Computers in Biology and Medicine*, accepted 2008.
2. S.-C. Lee and P. Bajcsy, "Trajectory fusion for three-dimensional volume reconstruction." *Computer Vision and Image Understanding* In Press, doi:10.1016/j.cviu.2007.02.005, Available online (2007)
3. A.Y. Lin, Z.M. Ai, S.-C. Lee, P. Bajcsy, J. Pe'er, L. Leach, A.J. Maniotis and R. Folberg, "Comparing vasculogenic mimicry with endothelial cell-lined vessels: Techniques for 3D reconstruction and quantitative analysis of tissue components from archival paraffin blocks." *Applied Immunohistochemistry and Molecular Morphology* **15** p113-119 (2007)

4. P. Bajcsy, "An Overview of DNA Microarray Image Requirements for Automated Processing." *Computer Society of India Communications* **30** p18-23 (2007).
5. S.-C. Lee and P. Bajcsy, "Understanding Challenges in Preserving and Reconstructing Computer-Assisted Medical Decision Process", *Proceedings of the 6<sup>th</sup> International Conference on Machine Learning and Applications, ICMLA 2007*, Cincinnati, USA, 2007.[submitted]
6. S.-C. Lee, P. Bajcsy, A.Y. Lin and R. Folberg, "Accuracy evaluation for region centroid-based registration of fluorescent CLSM imagery." *EURASIP Journal on Applied Signal Processing* **15** Article ID 82480, p1-11 (2006).
7. P. Bajcsy, "An overview of DNA microarray grid alignment and foreground separation approaches." *EURASIP Journal on Applied Signal Processing* **15** Article ID 80163, p1-13 (2006).
8. S.-C. Lee and P. Bajcsy, "Intensity correction of fluorescent confocal laser scanning microscope images by mean-weight filtering." *Journal of Microscopy* **221** p122-136 (2006).
9. P. Bajcsy, S.-C. Lee, A.Y. Lin and R. Folberg, "Three-dimensional volume reconstruction of extracellular matrix proteins in uveal melanoma from fluorescent confocal laser scanning microscope images." *Journal of Microscopy* **221** p30-45 (2006).
10. S.-C. Lee and P. Bajcsy, "Spatial Intensity Correction of Fluorescent Confocal Laser Scanning Microscope Images", *Proceedings of the European Conference on Computer Vision, Workshop on Computer Vision Approaches to Medical Image Analysis (ECCV/CVAMIA 06)*, Graz, Austria, 2006, Lecture Notes in Computer Science **4241** p143-54 (2006)
11. S.-C. Lee and P. Bajcsy, "Three-dimensional Volume Reconstruction Based on Trajectory Fusion from Confocal Laser Scanning Microscope Images" *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 06)* , New York, **Vol. 2**, p2221-2228 (2006).
12. S.-C. Lee and P. Bajcsy, "Automated Feature-based Alignment for 3D Volume Reconstruction of CLSM Imagery", *Proceedings of the SPIE International Symposium in Medical Imaging*, San Diego, p6144-105 (2006)
13. P. Bajcsy, S.-C. Lee and D. Clutter, "Supporting Registration Decision during 3D Medical Volume Reconstruction", *Proceedings of the SPIE International Symposium in Medical Imaging*, San Diego, p6144-119 (2006).
14. D.J. Scherba and P. Bajcsy, "Depth map calibration by stereo and wireless sensor network fusion." *Proceedings of the 8<sup>th</sup> International Conference on Information Fusion (FUSION)*, p8, July 25-28, 2005, Philadelphia, Pennsylvania.
15. M. Urban and P. Bajcsy, "Fusion of voice, gesture, and human-computer interface controls for remotely operated robot." *Proceedings of the 8<sup>th</sup> International Conference on Information Fusion (FUSION)*, p8, July 25-28, 2005, Philadelphia, Pennsylvania.
16. Peter Bajcsy, "An Overview of Microarray Image Processing Requirements.", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 20-26 June 2005, San Diego, CA.

#### Synergistic Activities.

- Reviewer for journals, for example: Pattern Recognition 2007, Journal of Microscopy 2006; Computing in Science & Engineering, 2006; IEEE Transactions on Geoscience and Remote Sensing, 2005; International Journal of Applied Math and Computer Science 2005; IEEE on Pattern Analysis and Machine Intelligence, 2004
- Served on the NIST advisory board for the SHIELD project 2003-2004
- Served on NSF Information Technology Research (ITR) Review Panels for Division of Informative Biology and Division of Information & Intelligent Systems, 2004 (twice) and 2005 (once),
- Served as a program committee member on the IEEE Workshop on Computer Vision Methods for Bioinformatics, San Diego, June 2005 (in conjunctions with IEEE CVPR 2005).
- Contributions to the science of learning; Collaborates with biologists, neuroscientists, psychologists and veterinary medicine experts on bio-computing problems.

#### Collaborators

Robert Folberg, Dept of Pathology, University of Illinois, Chicago, IL;  
 Praveen Kumar, Department of Civil and Environmental Engineering, UIUC, IL.  
 Barbara Minsker, Department of Civil and Environmental Engineering, UIUC, IL.  
 Gerald Nelson, Department of Agricultural and Consumer Economics, UIUC, IL;  
 Dan Kuchma, Department of Civil and Environmental Engineering, UIUC, IL.  
 Momcilo Markus, Illinois State Water Survey, Champaign, IL;  
 Jean-Christopher Lementec CHI Systems, Inc., Philadelphia, PA;  
 Wei Xie, ACNielsen, Chicago, IL;  
 Evan DeLucia, Dept of Plant Biology, University of Illinois, Urbana, IL;

#### **ALAN B. CRAIG**

National Center for Supercomputing Applications (NCSA)  
 University of Illinois at Urbana- Champaign  
 (217) 244-1988  
 acraig@ncsa.uiuc.edu

#### **Professional Preparation**

University of Illinois, Urbana, IL Secondary Education, B.S. 1983  
 University of Illinois, Urbana, IL Computer Science, M.S. 1985  
 University of Illinois, Urbana, IL Information, Science Ph.D. 2005

#### **Appointments**

University of Illinois at Urbana-Champaign:  
     National Center for Supercomputing Applications – UIUC 1987 - present  
     Visualization and Virtual Reality Group  
 Texas Instruments – Dallas, Texas 1985 – 1987  
     Defense Systems Electronics Group  
     Software Design Engineer – Computer Systems Training Group

#### **Publications (selected)**

- "Chapter 35 -- Scientific Visualization" Sherman, William R., Craig, Alan B., Baker, M. Pauline, Bushell, Colleen The Computer Science and Engineering Handbook Ed: Allen B. Tucker, Jr. CRC Press (Boca Raton, FL), 1997
- C. Scaletti, A. Craig, "Using Sound To Extract Meaning From Complex Data", Extracting Meaning from Complex Data: Processing, Display, Interaction II, Edward J. Farrell, Editor, Proc. SPIE 1459, 207-219 (1991)
- D. Kovacic, A. Craig, R. Patterson, W. Romme, D. Despain, "Fire Dynamics in the Yellowstone Landscape, 1690-1990: An Animation", Model Driven Visual Simulation, Proc. Resource

Technology 90, Second International Symposium on Advanced Technology in Natural Resources Management (1990)

W. Cheng, A. Craig, I. Dilber "Computer Simulation of Dynamic Impact Processes: A Visual and Audio Representation", Finite Element Applications, ASME Computers in Engineering Vol II (1991)

W. Cheng, A. Craig, I. Dilber "Impact Penetration Modeling Using DYNA3D" (Manuscript) 1991  
DYNA3D User Group Meeting, Bournemouth, England (1991)

**Books:**

Understanding Virtual Reality by William Sherman and Alan Craig – Morgan Kaufmann Publishing, September, 2002

Using Virtual Reality by Alan Craig, William Sherman, and Jeff Will – Morgan Kaufmann Publishing, in review

**Book Chapter:**

Chapter 194 – "Virtual Reality" Sherman, William R., Craig, Alan B. Encyclopedia of Information Systems Ed: Hossein Bidgoli Academic Press (San Diego, CA) 2003

**Articles:**

Literacy in Virtual Reality: A New Medium, William R. Sherman and Alan B. Craig, *Computer Graphics*, November 1995.

Virtual Reality Application Design Considerations Sherman, William R., & Craig, Alan B. Royal Society workshop on Virtual Reality in Society, Engineering and Science London, England, July 1995.

**Synergistic Activities**

Served as PI for the University of Illinois sub-award for NSF Grant # 0311088 "Scientific Visualization for Undergraduate Education."

Organized and hosted large-scale summer institute on Scientific Visualization (NSF Funded)

Organized and hosted multiple large-scale summer institutes on high performance computing  
Project leader for the NCSA VIAS information management system. The VIAS System

(Visualization Information Archival System) is an automated, webcrawling facility that builds databases on topics of interest. The first database the system built was for information regarding scientific visualization.

Have served as paper reviewer and on technical program committees for various conferences and journals

**Collaborators & Other Affiliations**

**• Collaborators and Co-Editors**

Eric W. Johnson – Valparaiso University

William R. Sherman – Desert Research Institute

Douglas Tougaw – Valparaiso University

Jeffrey D. Will – Valparaiso University

**• Graduate and Postdoctoral Advisors.**

Graduate Advisor – Tim Wentling, University of Illinois at Urbana-Champaign

**ALEX YAHJA**

4038 NCSA Building, M/C 257

National Center for Supercomputing Applications

1205 West Clark St.

Urbana, IL 61801

Education:

Carnegie Mellon University, M.Sc., 2000, (Robotics)  
Carnegie Mellon University, M.Sc., 2004, (Engineering and Public Policy)  
Carnegie Mellon University, Ph.D., 2006, (Computation, Organizations and Society)

Professional Experience:

National Center for Supercomputing Applications, 2006-present  
Assistant Director in Modeling, Institute for Computing in the Humanities, Arts, and Social  
Science

Publications (selected):

Alex Yahja, Systematic Modeling and Evaluation of Historic Processes, American Association for  
History and Computing, 2007, invited

Alex Yahja and Kathleen M. Carley, Simulation Validation: An Inferential Approach, Computational and  
Mathematical Organization Theory, 2007, accepted

Mengxiao Zhu, Alex Yahja and Noshir Contractor, The Investigation and Design of Socially Realistic  
First Responder Networks and Plans, the 4th UCLA Lake Arrowhead Conference on Human Complex  
Systems, Lake Arrowhead, CA, 2007.

Alex Yahja and Kathleen M. Carley, WIZER: Automated Model Improvement in Multi-Agent Social-  
Network Systems, in Scerri, P., Vincent, R., and Mailler, R.T. (Eds.), Coordination of Large-Scale  
Multiagent Systems VIII, Springer Verlag, 2005, ISBN: 0-387-26193-1

Alex Yahja and Kathleen M. Carley, An Inferential Approach to Validating Agent Simulations, Agent  
2007 Conference, Argonne National Lab, Chicago, IL, November 15-17, 2007

Alex Yahja, Sanjiv Singh, and Anthony Stentz, An Efficient On-line Path Planner for Outdoor Mobile  
Robots Operating in Vast Environments, Robotics and Autonomous Systems, Vol. 33, No. 2&3, August  
2000, pp. 129-143.

Research and Professional Record

North American Association for Computational Social and Organizational Science  
International Network for Social Network Analysis  
Society for Computer Simulations International

Research Interests

Human and social dynamics, computational social science, supercomputing, social networks, modeling  
and simulation, causal analysis, collaboration and recommender systems, computational organization  
theory, semantics & knowledge-based systems, and machine learning

Collaborators not mentioned in the publications

Elizabeth Casman, Douglas Fridsma, Demian Nave, Boris Kaminsky, Neal Altman, Li-Chiou Chen,  
Virginia Bedford

Ph.D. Advisor  
Kathleen Carley, Carnegie Mellon University

### **JAMES C. ONDERDONK**

University of Illinois at Urbana-Champaign  
Office of Continuing Education  
Conferences and Institutes  
302 East John Street, Suite 202  
Champaign, IL 61820  
Phone: (217) 333-2880; Fax: (217) 333-9561

1808 Floral Park Drive  
Savoy, Illinois 61874  
Phone (217) 359-5377

E-mail: onderdon@uiuc.edu

#### Education

Management of Lifelong Education, Harvard University, 1997  
Ph.D., Urban Services, Old Dominion University, Norfolk, Virginia, 1995  
M.S. Ed., Old Dominion University, 1977  
B.A., History, The College of William and Mary, Williamsburg, Virginia 1969

#### Professional Experience

Associate Director for Education and Outreach, Institute for Computing in Humanities, Arts, and Social Science, (I-CHASS), National Center for Supercomputing Applications, University of Illinois at Urbana- Champaign, Jan. 2007 – present

Head, Division of Conferences and Institutes, Office of Continuing Education, 1999-present and Interim Head and Associate Head, 1997-1999

In collaboration with members of the University community, Conferences and Institutes designs, plans and presents non-credit programs including conferences, seminars, short courses, certificate programs and annual meetings of professional societies. As Head, I:

- Provide day-to-day leadership of programmatic, financial, personnel and administrative functions of the division
- Represent the division to faculty, academic units and constituent communities to develop new outreach activities
- Supervise a staff of 15
- Prepare and monitor the divisional budget (\$2.9 million in FY07)
- Develop policies for appropriate expenditure of funds, cost recovery, and reconciliation of fund balances
- Provide liaison with the Illinois Manufacturing Extension Center
- Manage the Technical Assistance Center, a technology transfer program with the US Army's Construction Engineering Research Laboratory
- Act as administrative agent for Levis Faculty Center, a 20,000 sq. foot campus conference center

Adjunct Faculty, School of Education, Capella University, 2003-present

Teach online graduate courses in the history of higher education and finance of higher education.

Education Services Specialist, Federal Civil Service, Naval Medical Center, Portsmouth, Virginia. 1993-1997

- Managed all voluntary educational programs for personnel assigned to the Naval Medical Center, the largest center on the East Coast. Also responsible for programs at four other sites in southeastern Virginia.
- Planned and scheduled courses and degree completion programs -- remedial, vocational, undergraduate and graduate (78 courses with 1106 enrollments in

FY 96)

- Provided academic and vocational counseling and student services.
- Completed Contracting Officer's Technical Representative (COTR) training for contracts at Naval Education and Training Support Center, Atlantic.

Education Services Officer, Department of the Navy. 1982-1993

- Progressively more responsible assignments for Naval Education and Training Commands both in Norfolk, Virginia, and in the United Kingdom (1985-1989; managed seven different sites in Scotland, England and Wales)
- Supervised professional and support staff at the Navy's largest single education office in Norfolk, Virginia; secret security clearance.

Assistant P.A.C.E. Coordinator, Florida Junior College at Jacksonville, Norfolk, Virginia, Office. 1980-1982

- Coordinated the Program for Afloat College Education, a non-traditional higher education program providing educational opportunities for shipboard personnel.

Project Coordinator, Center for Educational Research, Old Dominion University.  
1979-1980

Assistant Instructor (Old Dominion University) and Adjunct Instructor (Christopher Newport College, Office of Continuing Education). 1975-1977

Reading consultant and classroom teacher, Newport News Public Schools, Newport News, Virginia.  
1970-1973 and 1977-1979

### Publications

Onderdonk, J. (2008, in press). Revivals in Antebellum America. In Gale Library of Daily Life: American Civil War. Farmington Hills, Michigan: Gale-Cengage.

Onderdonk, J. (2008, in press). Public Reaction to Lincoln's Gettysburg Address. In Gale Library of Daily Life: American Civil War. Farmington Hills, Michigan: Gale-Cengage.

Onderdonk, J. (2008, in press). The Freedom Riders: Two Personal Perspectives. In D. O'Brien (Ed.), Remembering Brown at Fifty: The University of Illinois Commemorates Brown v. Board of Education, Urbana: University of Illinois Press.

Burton, O., Onderdonk, J. & Appleford, S., (2007, in review). Keeping Up With the e-Joneses: The Role of Technology in the Democratization of Education. In W.Cope and M. Kalantzis (Eds.), Ubiquitous Learning, Urbana: University of Illinois Press.

Onderdonk, J., (2007). David Walker's Appeal. In Gale Library of Daily Life: Slavery in America. Farmington Hills, Michigan.

Onderdonk, J., (2007). The Confessions of Nat Turner. In Gale Library of Daily Life: Slavery in America. Farmington Hills, Michigan.

Burton, V., Appleford, S., & Onderdonk, J. (2007) A Question of Centers: One Approach to Establishing a Cyberinfrastructure for the Humanities, Arts, and Social Sciences. CTWatch (Cyberinfrastructure Technology Watch) Quarterly, Vol 3., No 2.  
<http://www.ctwatch.org/quarterly/articles/2007/05/a-question-of-centers/>

Onderdonk, J., Osteen, J. , & Brinton, R., (1996). PRIMEd to Break Ranks: Restructuring in the Norfolk School System. NASSP Bulletin (National Association of Secondary School Principals) 80, 82-89.

Onderdonk, J.C. (1995). Teacher Roles for the 21st Century. Unpublished doctoral dissertation, Old Dominion University, Norfolk, Virginia.

Presented

Onderdonk, J. (2007, October). RiverWeb: Technology and Outreach.  
Paper

Papers presented at the 2007 Outreach Scholarship Conference, Madison, WI.

Onderdonk, J. (2006, April). Evaluating Collaborative Programs – the Anna Karenina Approach.  
Paper presented at the University Continuing Education Association (UCEA) 91st Annual Conference,

San Diego, California  
Schejbal, D. & Onderdonk, J. (2000, January). To Collaborate or Not to Collaborate.  
Paper presented at the University Continuing Education Association  
Workforce Development Forum, St. Petersburg Beach, Florida

## Appendix VI: Curriculum Vitae- Humanities Group Leaders

### CHERYL E. BALL

Department of English  
Campus Box 4240  
Illinois State University  
Normal, IL 61790-4240

Office: 309. 438. 3152  
cball@ilstu.edu

#### ACADEMIC APPOINTMENTS

Assistant Professor, English Department, Illinois State University. 2007–present.  
Visiting Scholar, English Department, The Ohio State University. June 2007.  
Assistant Professor, English Department, Utah State University. 2004–2007.  
PhD in Rhetoric & Technical Communication, Michigan Technological University, 2005; Dissertation: A New Media Reading Strategy. Committee: Anne Wysocki (chair), Cynthia Selfe, Diana George  
MFA in Poetry, Virginia Commonwealth University, 2000. Thesis: Dinner for One [Firstinteractive, electronic thesis at VCU]. Chair: T.R. Hummer  
BA in English/Creative Writing, Old Dominion University, 1996. Magna Cum Laude  
(2006–present). Kairos: A Journal of Rhetoric, Technology, Pedagogy. Co-Editor.  
(2001–2006). Kairos: Rhetoric, Technology, Pedagogy. Co-Editor, CoverWeb section.  
(2002–present). Computers & Composition Online. Editorial Board Member.  
(2003–04). Computers & Composition. Associate Editor.  
(2000–03). Computers & Composition. Assistant Editor.  
(2002). College Composition & Communication. Editorial Intern.  
(1994–00). Various editorial roles in corporate and academic settings. Ask for details.

Ball, Cheryl E., & Kalmbach, Jim. (Eds.). (under contract). RAW: Reading and writing new media. Hampton Press: Creskill, NJ.

DeWitt, Scott Lloyd, & Ball, Cheryl E. (forthcoming: 2008, Summer). Kairos: A Journal of Rhetoric, Technology, Pedagogy, 12(3). [Special issue: Manifestos!].

Ball, Cheryl E., & Hawk, Byron. (Eds.). (2006, September). Computers & Composition, 23(3). [Special issue: Sound in/as compositional space: A next step in multiliteracies].

Ball, Cheryl E., & Hawk, Byron. (Eds.). (2006, Fall). Computers & Composition Online. [Special issue: Sound in/as compositional space]. <<http://www.bgsu.edu/cconline/sound>>.

Ball, Cheryl E., & Moeller, Ryan. (2007). Reinventing the possibilities: Academic literacy and new media. Fibreculture. <<http://journal.fibreculture.org/issue10/>>.

Arola, Kristin L., & Ball, Cheryl E. (2007). A conversation: From “They call me doctor?!” to tenure. Computers & Composition Online. <<http://www.bgsu.edu/cconline/doctor/>>. [Invited].

Anderson, Daniel; Atkins, Anthony; Ball, Cheryl E.; Homicz Millar, Krista; Selfe, Cynthia; & Selfe, Richard. (2006). Integrating multimodality in composition curricula: Survey methodology and results from a CCCC Research Initiative grant. Composition Studies, 34(2).

Ball, Cheryl E. (2006). Designerly . readerly: Re-assessing multimodal and new media rubrics for writing studies. Convergences, 12, 393–412.

Ball, Cheryl E. (2006). Reading the text: A rhetoric of wow. *Kairos: Rhetoric, Technology, Pedagogy*, 10(2). <<http://kairos.technorhetoric.net/10.2/>>.

Ball, Cheryl E. (2004). Show, not tell: The value of new media scholarship. *Computers & Composition*, 21, 403–425.

Kuhn, Virginia, & Ball, Cheryl E. (in revision). Embrace and ambivalence: The academy's love-hate relationship with the digital. *College Composition and Communication*.

Moeller, Ryan; Cargile Cook, Kelli; & Ball, Cheryl E. (accepted for collection). Political economy and sustaining the unstable: New faculty and research in English studies. In Danielle DeVoss, Heidi McKee, & Richard Selfe (Eds.) *Technological ecologies & sustainability*. [Collection under review.]

Ball, Cheryl E., & Arola, Kristin L. (2005). *ix tech comm: visual exercises for technical communication*. Boston: Bedford/St. Martin's Press.

Ball, Cheryl E., & Arola, Kristin L. (2004). *ix: visual exercises*. Boston: Bedford/St. Martin's Press.

Ball, Cheryl E. (2004). *Picturing texts instructor's guide*. New York: W.W. Norton.

Ball, Cheryl E. (2004). *Picturing texts website*. New York: W.W. Norton.  
<<http://www.picturingtexts.com>>.

Ball, Cheryl E. (2008–09). Digital scholarship in the Humanities, Part 1: Authors' composition and revision processes of new media scholarship. New Faculty Initiative Grant, Illinois State University. \$3,500.

Ball, Cheryl E. (2007–08). Digital Media and Composition (DMAC) institute. Professional Development Travel Grant program of the Faculty Excellence Initiative Committee. Illinois State University. \$750.

Ball, Cheryl E., & Moeller, Ryan. (2006–07). *The Learning Suite: A collaborative, technology-rich environment to support writing/composition in a digital age*. \$86,000. USU Innovation Fund.

## SERVICE

(2008–09). Chair. CCCC Committee for Computers in Composition & Communication (7Cs). [appointed]

(2007–present). Member. Task Force on Digital Scholarship for Tenure and Promotion. (Working group from C&W Conference.) [invited]

(2006–07). Chair. CCCC Writing Program Certificate of Excellence Selection Committee. [appointed]

(2006–08). Co-Chair. CCCC Committee for Computers in Composition & Communication (7Cs). [appointed]

(2005–08). Member. NCTE Committee on Technical & Scientific Communication. [appointed]

(2003–06). Member. CCCC Committee for Computers in Composition & Communication. [appointed]

## JOHN BONNETT

1 Chantler Road

(905) 735-8316

Welland, ON

Fax: (905) 984-4849

L3C-4M6

Email: jbonnett@brocku.ca

Citizenship: Canadian/American

### **Education:**

- Doctor of Philosophy, Department of History, University of Ottawa, 2002.  
Thesis: "Communication, Complexity and Empire: The Systemic Thought of Harold Adams Innis." Supervised by: Chad Gaffield
- Master of Arts Degree, Department of History, University of Ottawa, 1992.  
Research Paper: "Henry L. Stimson, *Mentalité*, and the Decision to Use the Atomic Bomb Against Japan." Supervised by: Brian Villa.
- Bachelor of Arts, Western Washington University, *Cum Laude*, 1989.

### **Professional Experience:**

- Tier II Canada Research Chair in Digital Humanities, Department of History, Brock University, July 1, 2005 to present
- Research Officer, National Research Council of Canada, Institute for Information Technology, March 2002 to June 30, 2005
- Adjunct Professor, Department of History, University of New Brunswick, October 2002 to Present
- Visiting Researcher, National Research Council of Canada, Institute for Information Technology, January 1998 to December 2001

### **Grants, Scholarships, and Academic Honors:**

- NCSA Visiting Researcher Fellowship, National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign, June to August 2007
- SSHRC President's Grant – 2005 Amount: \$35,000. Purpose: Financial Support for *The Computer: The Once and Future Medium for the Social Sciences and the Humanities* Conference.
- Canada Foundation for Innovation Conference Grant – 2005
- Brock Humanities Research Institute – 2005
- Canada Foundation for Innovation Infrastructure Grant -- 2005
- Tier II Canada Research Chair – 2005.
- Social Science and Humanities Research Council Doctoral Fellowship – 1997-1998.

### **Publications (selected):**

“Charting a New Aesthetics for History: 3D, Scenarios and the Future of the Historian’s Craft.” In *L’histoire sociale / Social History*. 40(79): 169-208.

“Mediating the Past in 3D, and how Hieroglyphs get in the way: The *3D Virtual Buildings Project*.” In *Mind Technologies: Humanities Computing and the Canadian Academic Community*. (Calgary: University of Calgary Press, 2005).

“Following in Rabelais’ Footsteps: Immersive History and the 3D Virtual Buildings Project.” In *History and Computing*. 13(2): 107-150. 2001 (Published 2004).

"Versando nuova linfa in una vecchia disciplina: Impiego del 3D per insegnare e rappresentare il passato." [English Title: "Pouring New Wine into an Old Discipline: Using 3D to Teach and Represent the Past."] in *Storic@mente*. 2004. On-line journal available at: [http://www.storicamente.org/02\\_tecnostoria/strumenti/bonnett\\_ital.htm](http://www.storicamente.org/02_tecnostoria/strumenti/bonnett_ital.htm) [December 22, 2004].

“New Technologies, New Formalisms for Historians: *The 3D Virtual Buildings Project*.” In *Literary and Linguistic Computing*. 19(3): 273-287. September 2004.

“The Oral Tradition in 3D: Harold Innis, Information Visualisation and the 3D Historical Cities Project.” In *Text Technology*. 12(1): 1-8. 2003.

“Following in Rabelais’ Footsteps: Immersive History and the 3D Virtual Buildings Project.” In *Journal of the Association for History and Computing*. 6(2). September 2003. Available on-line at: <http://mcel.pacificu.edu/jahc/jahcvi2/articles/bonnett/bonnett.html> [June 15, 2007]

“The Oral Tradition in 3D: Harold Innis, Information Visualisation and the 3D Historical Cities Project.” In *Computing in the Humanities Working Papers*. A.20. September 2003. Available on-line at: <http://www.chass.utoronto.ca/epc/chwp/titles.html> [April 30, 2005]

Bringing Students to a Virtual Past: Teaching Ottawa History with the *3D Historical Cities Project*." in *Construire une capitale -- Ottawa -- Making A Capital*. Eds. Jeff Keshen and Nicole Saint-Onge. (Ottawa: University of Ottawa Press, 2001): 483-502.

“Abductive Reasoning, A-Life and the Historian’s Craft: One Scenario for the Future of History and Computing.” At *Digital Arts and Humanities Network Website*. (Hosted by King’s College, London) Available on-line at: [http://www.arts-humanities.net/blog/ian\\_anderson/370](http://www.arts-humanities.net/blog/ian_anderson/370) [September 2007]

“High-Performance Computing: An Agenda for the Social Sciences and the Humanities in Canada.” **(Article commissioned by the Social Sciences and Humanities Research Council of Canada)** Available on-line at:

*Digital Arts and Humanities Network Website.* (Hosted by King's College, London)

[http://www.arts-humanities.net/blog/ian\\_anderson/393](http://www.arts-humanities.net/blog/ian_anderson/393)

*SSHRC Website: Social Sciences and Humanities Research Council.*

[http://www.sshrc.ca/web/about/publications/computing\\_final\\_e.pdf](http://www.sshrc.ca/web/about/publications/computing_final_e.pdf) [January 2007]

#### Scholarly Activities and Service:

Conference Organizer: 2006 Annual Meeting of the Society for Digital Humanities. Congress of the Social Sciences and the Humanities. York University. Toronto, ON, May 29-31, 2006.

Conference Organizer Symposium: The Computer – The Once and Future Medium for the Social Sciences and the Humanities. Congress of the Social Sciences and the Humanities. York University. Toronto, ON, May 30, 2006.

Member, Jury, Interactive Media Section, *Digital Humanities Quarterly*, February 2006 to present

Member, Editorial Board, *Journal of Canadian Studies*, August 2005 to present

Chair, Canadian Committee on History and Computing, Canadian Historical Association

### HUGH DENARD

#### Formal Education

---

- |         |  |
|---------|--|
| 1993-97 | Ph.D. in Drama - 'Modern Versions of Greek Tragedies from Ireland'<br>Supervisor: Leslie Read, Department of Drama, University of Exeter. British Academy Studentship. |
| 1992-93 | M.A. in Ancient Drama and Society, University of Exeter<br>Supervisor: John Wilkins, Department of Classics, Exeter.   |
| 1988-92 | B.A. in Drama & Classical Civilizations, Trinity College Dublin<br>Supervisor: Steve Wilmer, Samuel Beckett Centre, TCD.   |

#### Academic Posts

---

- |         |   |
|---------|---|
| 2005-   | Lecturer, Centre for Computing the Humanities, King's College, London (KCL) |
| 2000-05 | Lecturer, School of Theatre Studies, Warwick                                |
| 1998-00 | Teaching & Research Fellow, School of Theatre Studies, Warwick              |
| 1998-98 | Part-time Lecturer, School of English, Trinity College Dublin               |
| 1994-97 | Part-time Lecturer, Department of Drama, Trinity College Dublin             |

#### Major Roles (Current)

---

- Associate Director, King's Visualisation Lab, KCL
- Associate Director & Manager, Eduserv "THEATRON 3" Project
- Co-Director, Leverhulme "Roman Villa of Oplontis" Project
- Co-Director, AHRC "The Body and Mask in Ancient Theatre Space" Project
- Co-Director, AHRC "Making Space" Project
- Director & Manager, JISC 3D Visualisation in the Arts Network
- Editor & Joint Coordinator, *The London Charter*

- Programme Organiser, MA in Digital Culture and Technology, KCL

#### Selected Publications

---

- Editor-in-Chief, 2001-07; Editor (Resources), 2008-, *Didaskalia* ([www.didaskalia.net](http://www.didaskalia.net)): peer-reviewed e-journal, research and teaching resources dedicated to ancient drama in performance. Collaborations with King's College, London; American Philological Association; JISC-funded ARCHES Project. Three development grants.
- *Living Theatre: Roman Theatricalism and the Domestic Sphere* Co-author: Richard Beacham (Yale University Press: forthcoming)
- Curator, "The Centre for Computing in the Humanities, King's College London: A Second Life Exhibition."
- "Towards a Consensus in 3d Visualisation" *AHDS Newsletter* Spring/Summer 2007, pp.4-5.
- "Lost Theatre and Performance Traditions in Greece and Italy" *Cambridge Companion to Greek and Roman Theatre* Eds. J. Michael Walton and Marianne McDonald (Cambridge University Press, 2007), 139-160.
- "An Introduction to the London Charter" Co-authors: Richard Beacham and Franco Niccolucci, in Ioannides, M. *et al.* (eds), *The e-volution of Information Communication and Technology in Cultural Heritage* Refereed Proceedings of VAST Conference, Cyprus, November 2006. (Archaeolingua, 2006), pp. 263-269.
- *The London Charter for the Use of 3-dimensional Visualisation in the Research and Communication of Cultural Heritage*. [www.londoncharter.org](http://www.londoncharter.org) Draft 1: 5 March 2006. Draft 1.1: 14 June 2006.
- "Roman Theatre and Frescos: Intermedial Research Through Applied Digital Visualisation Technologies" Co-author: Richard Beacham, with Martin Blazeby in Hal Thwaites (ed.) *Virtual Reality at Work in the 21<sup>st</sup> Century: Impact on Society*. Proceedings of the 11<sup>th</sup> International Conference on Virtual Systems and MultiMedia (VSMM) 2005.
- "Transforming Online Learning Paradigms" *Interactions* Vol.7 No.2 (2003)
- "'At the foot of Pompey's Statue': Reconceiving Rome's *Theatrum Lapidium*" in Alan Bowman and Michael Brady Eds. *Images and Artefacts of the Ancient World* (Oxford University Press, 2005) 69-76.
- "Performing the Past: the Virtual Revolution in Performance History" in K. Schlesinger (ed.) *Performing Arts Resources* Vol.24 (NY: Theatre Library Association, 2004) 54-70.
- "The Art of Re-membling: Some Greek Tragedies from Ireland" Proceedings of IX International Meeting on Ancient Greek Drama (Delphi: 2004) 153-167. Co-author: Steve Wilmer
- "The Pompey Project: Digital Research and Virtual Reconstruction of Rome's First Theatre" Co-authored with Richard Beacham. ACH/ALLC Proceeding *Journal of Computers and the Humanities* Vol.37 No.1 (2003) 129-140.
- Introduction and Appendix to Aeschylus *Complete Plays Volume 1* tran. Carl Mueller (Hanover NH: Smith and Kraus, 2002) 1-49; 283-300. Reviewed in *Performing Arts Journal* 75, Vol.25, No.3, Sept. 2003
- "Research Recreates Ancient Roman Virtual Reality with 21st-century 3-D Technology" co-authored with Richard Beacham for University of Warwick *Humanities Research Centre Bulletin* (June 2003)
- Preface to Aeschylus *Complete Plays Volume 2* tran. Carl Mueller (Hanover NH: Smith and Kraus, Nov. 2002) ix-xiii.
- "Virtuality and Performativity: Recreating Rome's Theatre of Pompey" *Performing Arts Journal* 70 (2002) 25-43. Article highlighted in the *Chronicle of Higher Education's* Daily Report in February 2001.
- "Virtual Archaeology: Reconceiving Rome's Theatre of Pompey" *British Academy Review* (July-December 2001) 22-23.

- Preface to Sophokles *The Complete Plays* tran. Carl Mueller (Hanover NH: Smith and Kraus, 2000) vii-viii.

### Teaching

---

- Programme Organiser and Personal Tutor, MA in Digital Culture and Technology
- Module Organiser and Tutor, Digital Culture and Technology (MA Core Module)
- Module Organiser and Tutor, 3D Visualisation in the Arts, Humanities and Cultural Heritage
- Current MA and PhD supervision topics: Historical Visualisation; Greek and Roman Drama and its Reception.
- Early adopter / innovator in IT-augmented curriculum design, research- and resource-based learning, and theatre-historical teaching methods.

### Selected Papers and Presentations 2005-8

---

- “Recent Developments in Humanities Visualisation” (Paper) and “Grand Challenges in Arts and Humanities Visualisation” (Workshop Chair) VizNET 2008: The 2nd Interdisciplinary Conference on Intersections of Visualization Practices and Techniques, Loughborough, May 2008.
- “Visualising the imagined spaces of Pompeian frescoes” Workshop on “Reconstructing Pompeian Interiors: Painting, Models and Architectural and Virtual Reality”, Stockholm, April 2008.
- “The London Charter” Open Digital Cultural Heritage Systems Conference, Rome, Feb. 2008.
- “Digital Humanities Island” Colloquium on Shared Virtual Environments, University of Pisa, Feb. 2008.
- “Parallel universes, shared worlds: visualisation in the arts and humanities” Long Room Hub, TCD, Feb. 2008.
- “THEATRON 3” JISC CETIS-Eduserv Event, London Knowledge Lab, Sept. 2007.
- Co-organizer, “Standards, Scientific Reliability and 3D Visualization” panel, X3D, Perugia, April 2007.
- “The London Charter” Digital Resources for the Humanities and Arts, Dartington, Sept. 2006.
- Facilitator, “Making 3D Visual Research Outcomes Transparent”, British Academy and Expert Seminar, Centre for Computing in the Humanities, 23-5 Feb. 2006.
- "Intermediality, Interdisciplinarity, Reliability: Reflections on a Visualisation-based Study of Pompeian Frescos" CHArt: Computers and the History of Art conference, British Academy, Nov. 2005.

PATRICK DUNAE

e-mail: [dunae@mala.ca](mailto:dunae@mala.ca)  
 telephone: (250) 380-1633

### Present position

Professor  
 Department of History  
 Malaspina University-College  
 900 Fifth Street  
 Nanaimo, British Columbia  
 V9R 5S5

Adjunct Associate Professor  
 Department of History

University of Victoria  
3800 Finnerty Road  
Victoria, British Columbia  
V8W 3P4

## **Degree**

Ph. D. History. University of Manchester, 1976.

## **Recent publications and presentations:**

*Nanaimo, 1891: An historical GIS of a harbour city* (2007)

Published online at:

<http://tree.mala.bc.ca/nanaimo1891gis/>

“Reconstructing a harbour city in the Pacific Northwest with GIS: Nanaimo in the 1890s,” presented to the Social Science History Association, Chicago, Illinois, 16 November 2007

<http://web.mala.ca/dunae/hgis/index.htm>

“GIS and History,” presented to the international Pacific Region ESRI Users’ Conference, Victoria, B.C., 2 October 2007 [invited paper].

“*An historical HGIS in the Pacific Northwest: Directions from Victoria, B.C., 1871-1901*,” presented to the Social Science History Association, Portland, Oregon, 3 November 2005.

“*Visualizing Victorian Cities with HGIS [Historical Geographical Information Systems]*,” presented to the international Association for Computers and the Humanities conference, Victoria, British Columbia, 16 June 2005.

“*Virtual Victoria: Presenting the Past with New Technologies*,” (with John Lutz, University of Victoria), presented to a joint session of the Canadian Historical Association and the Canadian Association of Geographers, University of Western Ontario, London, Ontario, 1 June 2005.

## **Web sites**

(2006) *viHistory.ca* Launched at Malaspina University-College in 2003 and re-developed with the Humanities Media and Computing Centre at the University of Victoria, this web site provides a searchable, digital archive of census data, directories, tax assessment rolls, and maps of Vancouver Island, c. 1861- 1911. The data set contains over 200,000 records and is used extensively by scholars, students and the general public.

<http://vihistory.ca>

(2005) *Virtual Victoria: View from the steeple, 1891*. This interactive web site demonstrates how archival material and electronic media can provide a compelling perspective on the past.

<http://cdhi.mala.bc.ca/steeple/>

## STUART DUNN

### EMPLOYMENT

- 2006 – present: Arts and Humanities e-Science Support Centre, King's College London:
  - Research Associate
- 2003 – 2006: AHRC ICT in Arts and Humanities Research Programme, University of Reading:
  - Programme Administrator (June 2005 – January 2006)
  - Research Assistant (November 2003 – June 2005)
- 2002 – 2003: Fenwick of Newcastle Ltd:
  - Various roles including analyst, training officer and section head
- 1999 – 2001
  - Course Tutor for 'The Art and Myth of Ancient Greece' and 'Minoans and Mycenaeans', University of Durham (1999 – 2001)
  - College Tutor, Trevelyan College, University of Durham (1999 – 2001)

### EDUCATION

- 2002: Phd (Durham) in Aegean Bronze Age archaeology (Thesis: *The chronology of the Aegean Late Bronze Age with special reference to the 'Minoan' eruption of Thera*)
- 1998: BA (II, First Division Joint Honours) in Ancient History and Archaeology, University of Durham

### RESEARCH AWARDS

2007:

- AHRC ICT Methods Network grant to convene 'Space and Time: methods in geospatial computing for mapping the past' (£4949.28).

### PROFESSIONAL ACTIVITIES

2007:

- Chair, EVA conference programme committee (London)
- Member, Location and the Web programme committee (Beijing)
- Workpackage Programme Manager, 'Enabling e-Uptake of e-Infrastructure Services'.
- Co-convenor, 'E-Science for the Arts and Humanities: an early adopters' forum'. Workshop held at NCSA, Urbana-Champaign, Illinois, June 2007.

2006:

- Evaluator, JISC Digitization Programme
- Visiting Research Fellow, School of Human and Environmental Sciences (Archaeology Department), University of Reading, UK (until 2009).
- UK e-Science All Hands Conference Programme Committee
- Member, Pleiades Project Technical Oversight Board, Ancient World Mapping Center, University of North Carolina Chapel Hill (invited member; ongoing).

2005 - 2006:

- DigitalClassicist project advisory group (invited member; ongoing)
- JISC Geospatial Data Workgroup (invited member; ongoing)

2004-present:

- Silchester Roman Town: A Virtual Research Environment for Archaeology, University of Reading (archaeological consultant and invited steering cttee member; ongoing)

#### SELECTED CONFERENCE PRESENTATIONS

2007:

- ‘The Anthropology of knowledge: from basic to complex communities in the Arts and Humanities’. Digital Humanities 2007, Urbana-Champaign, Illinois
- ‘Space as an artefact: understanding past perceptions and uses of space with and without computers’. Digital Classicist Seminar Series, KCL, August 2007.
- ‘A new way of working: the UK’s Arts and Humanities e-Science Initiative’. Acume2 project seminar, Warsaw, Poland.
- ‘A point in space, a moment in time: towards an integrated view of the Santorini eruption’. Minoan Chronology Workshop, Sonderborg, Denmark.

#### SELECTED PUBLICATIONS

2008:

- (with T. Blanke): Next Steps for E-Science, the Textual Humanities and VREs  
A Report on Text and Grid: Research Questions for the Humanities, Sciences and Industry, UK e-Science All Hands Meeting 2007. D-Lib Magazine Volume 14 Number 1/2 (January/February 2008): <http://www.dlib.org/dlib/january08/dunn/01dunn.html>

2007:

- (with Alex Voss, Matthew Mascord, Mercedes Arguello Castelleiro, Marzieh Asgari-Targhi, Rob Procter, Michael Fraser, Marina Jirotko, Peter Halfpenny, David Fergusson, Malcolm Atkinson, Tobias Blanke, Lorna Hughes and Sheila Anderson): e-Infrastructure development and community engagement. In *Proceedings of e-Social Science 2007*.
- Trustworthy Characters: common issues for archaeology, classical studies and VREs. Proceedings of First International Workshop on VREs, Edinburgh, March 2007
- Review: Georeferencing: The Geographic Associations of Information. *Literary and Linguistic Computing* Vol. 22 (2), June 2007: 243-245.
- (with A. Aschenbrenner, T. Blanke, M. Kerzel, A. Rapp and A. Zielinski): Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits – und Kreativbereich für Kunst und Kultur. *Bibliothek: Forschung und Praxis* 31 (1) 2007: 11-21

2006:

- (with Lorna Hughes and Nicolas Gold): CHIMERA: A serice-oriented approach to archaeological research. In *Proceedings of Computational Applications and Quantitative Methods in Archaeology, Berlin 2007*.
- (with Tobias Blanke): The UK Arts and Humanities e-Science Initiative. *Proceedings of IEEE conference, Amsterdam 2007*.

2005:

- (with Lorna Hughes and Sheila Anderson): Virtual Research Environments in the Arts and Humanities, *Proceedings of the e-Science All Hands Meeting 2005*,
- From Juktas to Thera: people and their environment in Middle and Late Minoan Crete. In A. Dakouri-Hild and S. Sherratt (eds.) *Autochthon: Papers presented to Oliver Dickinson on the occasion of his retirement*. BAR International Series, Oxford

2004:

- GIS and databases in Aegean prehistory: current practice, future strategy *Archaeological Computing Newsletter*, 61, December 2004

## JASON GILLILAND

Department of Geography, Social Science Centre  
The University of Western Ontario, London, ON Canada N6A 5C2  
Tel: (519) 661-2111x81239, [spatialists@gmail.com](mailto:spatialists@gmail.com)  
[www.spatialists.com](http://www.spatialists.com), [www.imagininglondon.ca](http://www.imagininglondon.ca)

### Academic Positions

2007-present Associate Professor, (tenured), Geography, University of Western Ontario  
2006-present Director, Urban Development Program, University of Western Ontario  
2006-present Associate Scientist, Children's Health Research Institute  
2003-2007 Assistant Professor (tenure track), Geography, University of Western Ontario  
2001-2003 Postdoctoral Fellow (SSHRC), Geography & Planning, U of Toronto

### Education

2001 PhD, Geography, McGill University, Montreal, Canada  
1996 M.Architecture, Architecture, McGill University, Montreal, Canada  
1994 M.A., Geography, McGill University, Montreal, Canada  
1991 B.A.(Hon), Geography, McMaster University, Hamilton, Canada

### Current Research Grants

Start-End Date	Position	Granting Agency	Grant Title	Total Amount
2007-08	principal investigator	CHRI	Geographical analysis of paediatric trauma in Southwestern Ontario	\$7,500
2007	principal investigator	Heart & Stroke Fdn	Development grant for project on Obesity & Built Environment	\$3,200
2007-08	principal investigator	LHSC – pediatrics	Geographic analyses of paediatric trauma	\$5,000
2007-08	principal investigator	LHSC – trauma	Geographic studies of paediatric trauma incidences in SW Ontario	\$5,000
2006-09	principal investigator	SSHRC	Social mobility in Canadian cities, 1880-1914	\$178,000
2006-09	co-investigator (p.i. J. Baxter)	SSHRC	Environmental inequity in Canada: patterns and experiences	\$112,100
2006-09	collaborator (p.i. J. Parr)	SSHRC	Lostscapes: New Media Reconstructions of Disrupted Environments	\$120,000
2005-08	principal investigator	CIHR	Environmental influences on obesity-related behaviours in youth	\$169,141

### Selected Publications

Larsen, K., and J. Gilliland. 2008. Mapping the evolution of 'food deserts' in a Canadian city: Supermarket accessibility in London, Ontario, 1961-2005. *International Journal of Health Geographics*

- Tucker, P., and J. Gilliland. 2007. The Effect of Season and Weather on Physical Activity: A Systematic Review. *Public Health* 121(12): 909-922.
- Tucker, P., Gilliland, J., Irwin, J. 2007. Splashpads, swings, and shade: parents' preferences for neighbourhood parks. *Canadian Journal of Public Health* 98(3): 198-202.
- Gilliland, J. and M. Novak. 2006. Positioning the past with the present: on integrating fire insurance plans and GIS for urban environmental history. *Environmental History* 11(1): 136-139.
- Gilliland, J. and P. Gauthier. 2006. The study of urban form in Canada. *Urban Morphology* 10(1): 51-66.
- Gauthier, P. and J. Gilliland. 2006. Mapping urban morphology: a classification scheme for interpreting contributions to the study of urban form. *Urban Morphology* 10(1): 41-50.
- Gilliland, J., Holmes, M., Irwin, J. and T. Tucker. 2006. Environmental equity is child's play: mapping recreational opportunities in urban neighbourhoods. *Vulnerable Children and Youth Studies* 1(2): 1-13.
- Wilson, D., Gilliland, J., Ross, N., Derevensky, J. and R. Gupta. 2006. Video Lottery Terminal Access and Gambling Among High School Students in Montréal *Canadian Journal of Public Health* 97(3): 202-206.
- Gilliland, J. and N. Ross. 2005. Opportunities for video lottery gambling: an environmental analysis. *Canadian Journal of Public Health* 96(1): 55-59.
- Gilliland, J. 2004. Muddy shore to modern port: the redimensioning of Montreal's waterfront time-space. *Canadian Geographer* 48(4): 448-472.
- Gilliland, J. 2003. It's About Time: Exploring the Fourth Dimension in the Morphology of Urban Disasters. *Urban Morphology* 7(2): 110-112.
- Gilliland, J. and S. Olson. 2003. Montréal, l'avenir du passé. *GÉOinfo*, (Jan-Feb): 5-7.
- Gilliland, J. 2002. The Creative Destruction of Montreal: Street Widening and Urban (Re)Development in the Nineteenth Century. *Urban History Review / Revue d'histoire urbaine* 31(1): 37-51.
- Gilliland, J. 2002. Society and Space in the Industrial City, *Urban History Review / Revue d'histoire urbaine*, 31(1): 3-4.
- Gilliland, J. 2000. Urban morphology in the 'Steel City'. *Urban Morphology*, 4(2): 102-3.
- Gilliland, J. 2000. Visions and Revisions of House and Home: A Half-Century of Change in Montreal's 'Cité-jardin'. In H. Nicol and G. Halseth (eds), *(Re)Development at the Urban Edges*. University of Waterloo, pp.139-74.
- Gilliland, J. 1999. Redimensioning the Urban Vascular System: Street Widening Operations in Montreal, 1850-1918. In G. Corona and G. L. Maffei (eds), *Transformations of Urban Form: From Interpretations to Methodologies in Practice*. (Firenze: Alinea Editrice) pp.FK2.7-10.
- Sendbeuhler, M. and J. Gilliland, 1998. "...to produce the highest type of manhood and womanhood": The Ontario Housing Act, 1919, and a New Suburban Ideal. *Urban History Review / Revue d'histoire urbaine* 26(2): 42-55.

Gilliland, J. and S. Olson, 1998. Claims on Housing Space in Nineteenth-Century Montreal. *Urban History Review / Revue d'histoire urbaine* 26(2):3-16.

Gilliland, J. 1998. Modeling Residential Mobility in Montreal, 1860-1900. *Historical Methods* 31(1): 27-42.

### GAIL E. HAWISHER

Department of English  
Illinois  
608 South Wright Street  
Urbana, IL 61801  
(217) 333-3251  
Fax 217-333-4321

603 West Church Street University of  
Savoy, Illinois 61874  
Telephone (217) 352-8031

hawisher@uiuc.edu

### EDUCATION

University of Illinois, Urbana-Champaign: Ph.D. in Composition Studies. January 1986.

The Ohio State University: Graduate work in Rhetoric, Literature, and Linguistics. 1977-1982.

Augusta College: B.A. in English. June 1970.

### EXPERIENCE (selected)

Professor of English and Director of the Center for Writing Studies. University of Illinois, Urbana, Illinois. July 1990-present. (promoted from Associate Professor in 1996)

Assistant Professor of English: Purdue University, West Lafayette, Indiana. 1989-1990.

Assistant Professor of English: Illinois State University, Normal, Illinois. 1986-1989.

### PUBLICATIONS (selected)

#### Books

Gaming Lives in the 21st Century: Literate Connections. (edited with Cynthia L. Selfe). New York: Palgrave Macmillan, 2007. (273 pages).

Literate Lives in the Information Age: Narratives of Literacy from the United States. (coauthor with Cynthia L. Selfe). Mahwah, NJ: Erlbaum, 2004. (259 pages).

Global Literacies and the World Wide Web. Ed. Gail E. Hawisher and Cynthia L. Selfe. New York: Routledge, 2000. (299 pages)

Passions, Pedagogies, and 21st Century Technologies. Ed. Gail E. Hawisher and Cynthia L. Selfe. Logan: Utah State University Press, 1999. (452 pages)

Literacy, Technology, and Society: Confronting the Issues. (edited with Cynthia Selfe). Upper Saddle River, NJ: Prentice Hall, 1997. (606 pages). (College Reader)

Computers and the Teaching of Writing in American Higher Education, 1979-1994: A History. (co-authored with Paul LeBlanc, Charles Moran, and Cynthia Selfe). Norwood, NJ: Ablex Publishing, 1996. (363 pages).

Re-Imagining Computers and Composition: Teaching and Research in the Virtual Age. (edited with Paul LeBlanc) Portsmouth, NH: Boynton/Cook, 1992. (222 pages)

Evolving Perspectives on Computers and Composition Studies: Questions for the 1990s. (edited with Cynthia L. Selfe) Urbana, IL: NCTE, 1991. (383 pages).

On Literacy and Its Teaching: Issues in English Education. (edited with Anna O. Söter) Albany, NY: SUNY Press, 1990. (259 pages).

Critical Perspectives on Computers and Composition Instruction. (edited with Cynthia Selfe) New York: Columbia University's Teachers College Press, 1989. (231 pages)

#### Articles

- “Globalization and Agency: Designing and Redesigning the Literacies of Cyberspace.” (with Cynthia Selfe, Yi-Huey Guo, and Lu Liu) *College English*. 68 (July 2006): 619-636.
- “Becoming Literate in the Information Age: Cultural Ecologies and the Literacies of Technology.” (with Cynthia Selfe, Brittney Moraski, Melissa Pearson). *College Composition and Communication*. 55.4 (June 2004): 642-92.
- “Collaborative Configurations: Researching the Literacies of Technology.” (with Cynthia Selfe). *Kairos*. 7.3 (Fall 2002):  
<http://english.ttu.edu/kairos/7.3/binder2.html?coverweb/hawisher/index.htm>.
- “A Historical Look at Electronic Literacy: Implications for the Education of Technical Communicators.” (with Cynthia Selfe). *Journal of Business and Technical Communication*. 16.3 (July 2002): 231-276.
- “Accessing the Virtual Worlds of Cyberspace.” *Journal of Electronic Publishing*. 6.1 (2000). University of Michigan. <http://www.press.umich.edu/jep/06-01/hawisher.html>.
- “Constructing Identities through Online Images.” *JAAL*. (March, 2000).  
[http://www.readingonline.org/electronic/jaal/3-00\\_Column.html](http://www.readingonline.org/electronic/jaal/3-00_Column.html)
- Rpt. in *Literacy in the Information Age: Inquiries into Meaning Making with New Technologies*. Bertram C. Bruce, ed. Newark, Delaware: International Reading Association, 2003. 128-140.
- “Reflections on Research in Computers and Composition Studies at the Century’s End.” (with Cynthia Selfe). *The Australian Journal of Language and Literacy*, 19.4 (1996): 290-304.  
 Rpt. as lead chapter in *Page to Screen: Taking Literacy Into the Electronic Era*. Ilana Snyder, ed. New South Wales, Australia: Allen and Unwin, 1997. 3-19.  
 Rpt. in *Teaching Literacy Using Information Technology*. Joeli Hancock, ed. Victoria: Australian Literacy Association, 1999. 31-47.
- “Writing Across the Curriculum Encounters Asynchronous Learning Networks or WAC Meets Up with ALN.” (with Michael Pemberton) *Journal of Asynchronous Learning Networks*. 1.1 (1997). [http://www.aln.org/alnweb/journal/jaln\\_issue1.htm#hawisher](http://www.aln.org/alnweb/journal/jaln_issue1.htm#hawisher)  
 Rpt. in *Communication Across the Curriculum*. Donna Reiss, Richard Selfe, and Art Young, eds. Urbana, IL: National Council of Teachers of English, 1998. 17-39.  
 Rpt. in *A Guide to Online course Development: The Theory and Practice of Online*

#### GRANTS, FELLOWSHIPS, AWARDS (selected)

- Outstanding Technology Innovator Award presented by the Conference on College Composition and Communication, May 2000.
- Distinguished Book Award for Passions, Pedagogies, and 21st Century Technologies (with Cynthia Selfe) presented at the 2000 Computers and Writing Conference. Texas Women’s University. Fort Worth, Texas.
- “Technological Literacy in America: Tracing the Paths of the Technology-Linkage.” (with Cynthia Selfe). National Council of Teachers of English Research Foundation Grant. Urbana, IL. 2000. \$12,400.00. (funded)
- “Technological Literacy in America, 1978-2000: A Project to Improve Technical Communication Education in the New Millennium.” (with Cynthia Selfe). Special Opportunities Grant, Society for Technical Communication. 2000. \$10,000.00. (funded)

**GERALDINE HENG**  
 Director, Medieval Studies  
 University of Texas at Austin

University of Texas at Austin  
Parlin 108, Austin, TX 78712-1164  
Tel. No. (512) 471-7480 (office)  
Fax: (512) 471-4909 (office)

Austin, TX 78722  
Tel. (512) 472-5472  
e-mail: heng@mail.utexas.edu  
Fax: (512) 472-5472 (home)

## ***EDUCATION***

Cornell University	PhD in English, 1990
	MA in English, 1986
National University of Singapore	MA in English, 1980
University of Singapore	BA (Honors, English and Philosophy), 1979

## **PUBLICATIONS** (selected publications only)

Book: Empire of Magic: Medieval Romance and the Politics of Cultural Fantasy. NY: Columbia University Press, 2003. 521 + xii pp. (Finalist, First Book Award, Modern Language Association of America, 2004; 2004 Best Book Award, SCMLA)

Edited anthology: Medieval Race. Medieval Academy of America MART critical scholarship series Vol.3, University of Toronto Press, 2008.

“The Race Template.” Cloning Cultures. Ed. Philomena Essed. Duke University Press, forthcoming, 2008.

“Jews, Saracens, “Black men,” Tartars: England in a World of Racial Difference, 13th-15<sup>th</sup> Centuries.” A Companion to Medieval English Literature, c.1350-c.1500. Ed. Peter Brown. Blackwell Companions to Literature and Culture. Blackwell, 2007. 247-269.

“An Experiment in Collaborative Humanities: ‘Global Interconnections: Imagining the World 500-1500.’” ADFL Bulletin, Modern Language Association of America, 38:3, Dec 2007.

“Pleasure, Resistance, and a Feminist Aesthetics of Reading.” Cambridge Companion to Feminist Literary Theory. Ed. Ellen Rooney. Cambridge UP, 2006. 53-72.

“Global Interconnections: Imagining the World 500-1500.” Medieval Academy Newsletter, September 2004.

“Race Returns West to Medieval Mindset.” The Times Higher Education Supplement. Dec 20, 2002.

“The Romance of England: Richard Coer de Lyon, Saracens, Jews, and the Politics of Race & Nation.” The Postcolonial Middle Ages. Ed. Jeffrey Cohen. The New Middle Ages Series. NY: St Martins, 2000. 135-171.

## **Selected presentations and Workshops:**

Lecture and workshop: “The Global Middle Ages Project and the Mappamundi Digital Initiative.” University of California Humanities Research Institute, Irvine, California, June 5, 2007.

Co-organizer, first planning workshop of the Global Middle Ages Project (G-MAP) and the Mappamundi online digital initiative, University of Minnesota, November 8-11, 2007.

Organizer/chair (by invitation), “Mappamundi: A Global Middle Ages.” International Congress of the New Chaucer Society, July 2006.

“An Experiment in Collaborative Humanities: ‘Global Interconnections: Imagining the World 500-1500.’” Presentation and Workshop on the Global Middle Ages Project and the Mappamundi Digital Project, Humanities Research Institute, University of California at Irvine, June 6, 2007.

Invited Lecture: “‘Global Interconnections’ and the Mappa Mundi Project.” Cornell Club of New York, 2008.

“An Experiment in Collaborative Humanities: ‘Global Interconnections: Imagining the World 500-1500.’” Symposium on New Directions in the Humanities. Columbia University, February 22-23, 2007.

## SERVICE

### Professional Service: National

Founder and Co-Director, The Global Middle Ages Project (G-MAP) and the Mappamundi Digital Online Initiative: multi-campus, interdisciplinary projects in the Humanities (in collaboration with several institutional partners), 2007-

Executive Committee, Comparative Medieval Literature Division, Modern Language Association (MLA), 2008-12

Chair, Executive Committee, Middle English Division, Modern Language Association (MLA), 2003-4

Secretary, Executive Committee, Middle English Division, Modern Language Association (MLA), 2002-3

Executive Committee, Middle English Division, Modern Language Association (MLA), 2000-2005

Steering committee on international feminism, National Women's Studies Association, 1999

Elected Delegate, Delegate Assembly, Modern Language Association (MLA), 1993-1995

Reader, PMLA, Speculum, Journal of Medieval & Early Modern Studies, Journal of English & Germanic Philology, Studies in the Age of Chaucer, Feminist Studies, differences, Texas Studies in Language & Literature, Stanford Humanities Center, university presses (e.g. Harvard, Duke, Indiana, Florida, Brepols, Palgrave, Macmillan)

Editorial Advisor, Cursor Mundi: Viator Studies of the Medieval and Early Modern World, UCLA Center for Medieval and Renaissance Studies

### **College & University Service**

Director, Medieval Studies Program (development of new interdisciplinary, collaborative graduate seminars under the rubric, “The Global Middle Ages” and “Medieval Cultural Studies;” the graduate Papers-in-Progress series; the Distinguished Visiting Lecturers series; interdisciplinary graduate symposia and professionalization workshops; lectures, presentations and public events at Explore UT, annual university open house; public outreach and community directives), 2002-5

Chair, Graduate Studies Committee, and Advisor, Medieval Studies Program, 2002-5

Associate Director and Co-founder, The Humanities Institute (conceptualized and initiated the Humanities Institute: the multi-college, interdisciplinary Humanities Seminar, and the Distinguished Lecturers in the Humanities Series), College of Liberal Arts, 2001-2

### **Virginia Kuhn, PhD**

2100 Griffith Park Blvd. #6 Los Angeles, CA 90039

213.924.7315 email: vkuhn@cinema.usc.edu

### **Employment:**

**2007- present University of Southern California (USC), Associate Director**

**Director of Honors in Multimedia Scholarship Program**, Institute for Multimedia Literacy (IML)

**Research Assistant Professor**, School of Cinematic Arts, USC

**2006-7 Project Specialist**, IML, USC

**2005-6 Postdoctoral Research Associate**, IML, USC.

Dissertation: *Ways of Composing: Visual Literacy in the Digital Age.* Director: Alice Gillam

PhD: English, Rhetoric & Composition University of Wisconsin, Milwaukee August 2005

MA: English, Rhetoric & Composition UWM 1999

BA: Art History and Criticism, UWM, 1984. Liberal Arts Curriculum Arizona State University, Tempe Arizona, 1979 - 82.

**Publications:**

- "From Gallery to Webtext: A Multimodal Anthology." Kairos, Spring 08.
- "Representing Lives: Whose Pictures are Worth 1000 Words? Image Events: From Theory to Action. Forthcoming edited anthology.
- "Cyber-Interface: The Foreign Techno-Human Interaction," The Structure of Cyberinfrastructure; Networked Knowledge Technologies and the Humanities and Interpretative Social Sciences. Forthcoming edited anthology.
- "TK3: A Tool to (Re)Compose," Academic Commons, December, 2005.
- "Picturing Work: Visual Projects in the Writing Classroom," Kairos 9.2, Spring 2005.
- "The Fruits of Our Labor," Conference Review of the 2004 CCCC.
- "Review: Virtual Education, Real Educators." Kairos 8.2, Winter, 2004.
- "The Rules of the Game," Book Review of Cybertext: Yearbook 2000," ebr, Sept. 2002.

**Invited Speaking Engagements:**

- HPC Horizons Inaugural Conference, March 13, 2008.
- USC's Center for Excellence in Teaching, November 14, 2007.
- UCLA Digital Humanities Center, November 7, 2007.
- The Academy's Love-Hate Affair: Resistance to Digital Scholarship, City University of New York's Graduate Program. October 2006.
- New Digital Tools: Sophie. Faculty training & Course lecture. Conceptual Studies in Media Arts Production and Cultures & Communities, University of Wisconsin, Milwaukee. October 2006.
- The Digital Portfolio Project: Cyberinfrastructure Institute, San Diego Supercomputing Center, July 2006.
- Transformation Through Use: The Pedagogy of Digital Writing Institute of Multimedia Literacy, University of Southern California, September 2005.
- Theory/Practice Colloquia: Conceptual Studies in Media Arts Production, Department of Film, University of Wisconsin, Milwaukee, September, 2005.

**Academic Presentations:**

- Designing Our Web 3.0 Lives: The Virtual is the Real. Society for Cinema and Media Studies Conference, 2008.
- New Media Scholarship Stakeholders: Departmental, Authorial and Editorial Issues. Computers & Writing Conference, 2008.
- You and YouTube: The Pedagogy of Production, CCCC 2008.
- "Is the Personal Still the Political? Visual Identity in the Multicultural Classroom." International Visual Literacy Association Conference. 2006.
- "From Panel to Gallery: Twelve Exhibits: Digital Illumination." CCC 2006
- "Reluctant Collaborators: The Challenges of the Student-Centered Classroom. AILA July, 2005
- "Multiculturalism & 21st Century Tool Kits" Computers & Writing, 2005.
- "Tools of Design" CCCC, 2005.
- "Virtual Selves: The Ebb & Flow of Online Identity," Computers and Writing Conference, 2004. "New Vistas for Rhetorical Inquiry: From the Ancients to the Internet," Conference of the Rhetoric Society of America, 2004.
- "You're So Paranoid I Bet You Think This Panel's About You: Rooting out Paranoia in the Profession," CCCC 2004.
- "How Do We Teach the Teachers?" New Teacher Preparation Forum, Computers and Writing Conference, 2003.
- "Performing Stereotypes." Midwest Conference of Language, Literature and Media.2003.
- "Reworking Identities in Writing Program Administration: Working Within the Gaps of Administrator, Faculty & Student." Watson Conference on Rhetoric and Composition 2002.

"The Politics of Grammarcheck." Computers & Writing Conference, 2002.  
 "Enhanced Teaching Through Electronic Listening: Reflections on the Digital Composition Classroom," Center for Professional Development Symposium, University of Wisconsin, Milwaukee. 2001.  
 "Using Literature in the Writing Class," Convention of the NCTE. 2000.  
 "Teaching With Technology: Programs, Pitfalls, Potential," English Department Teacher Orientation, UWM 2001.  
 "Face to Face and E-space: The Dynamics of the Computer Mediated Classroom," Conference on College Composition and Communication, 2000.  
 "Online Dramatic Interchange" Those Who Can Teach. Sixth Annual Conference on First Year Composition, 1999.

**Grants & Awards:**

Data Allocation on Tera-Grid for the Digital Portfolio Project. San Diego Supercomputing Center. 2007-08.  
 Recipient of the UWM Chancellor's Scholars' Award for Digital Scholarship. 2004-05  
 Writing Program Administration Team: Preparing Future Faculty competitive grant, UWM, 2001-02.  
 "The Digital Composition Classroom," Preparing Future Faculty competitive grant, University Of Wisconsin, Milwaukee, 2000-2001.  
 Fellowship, Illinois State University. Fall 1999 term.

**SUSAN J. NOAKES**

Department of French and Italian 260 Folwell Hall University of Minnesota Minneapolis, MN 55455 (612) 624-0076	Center for Medieval Studies 301 Nolte Center University of Minnesota Minneapolis, MN 55455 (612) 625-3034	436 Ashland Avenue St. Paul MN 55102 (651) 229-0295
--	---	---

**EDUCATION (selected):**

University of Chicago, A.B. with honors, 1967  
 Cornell University, Comparative Literature, 1969  
 Yale University, Comparative Literature, M. Phil., 1972, Ph.D., 1975

**ACADEMIC APPOINTMENTS:**

Yale University: Acting Instructor, 1973  
 University of Chicago: Asst. Prof. of Comparative Literature, Romance Languages, and General Studies in the Humanities, 1974-81; Associated Faculty Member, Graduate Library School, 1975-81; Instructor, "Basic Program" (Liberal Arts for Adults), 1980-81  
 University of Kansas: Asst. Prof. of French and Italian, 1981-83; Assoc. Prof. of French and Italian, 1983-88; Director, Women's Studies Program, 1986-88  
 University of Minnesota: Prof. of French and Italian, 1988-; Interim Director, Center for Advanced Feminist Studies, 1991; Associate Dean for Faculty, College of Liberal Arts, 1991-94; Director, Center for Medieval Studies, 2002-  
 University of North Carolina, Chapel Hill: Visiting Professor of Women's Studies, Fall, 1995, 1996

**PROFESSIONAL SERVICE:**

*Editorial Positions:*

Medieval Encounters: Jewish, Christian and Muslim Culture in Confluence and Dialogue, co-editor, 2005-2008, lead co-editor 2005.  
Dante Studies, Journal of the Dante Society of America (Associate Editor) 1993-  
Speculum, Journal of the Medieval Academy of America,

Book Review Editor, Italian studies, 2005-2006

Gradiva, 1980-85; International Courtly Literature Monograph Series, 1978

Signs (Associate Editor), 1990-91

Consultant Reader

Presses

Univ. of Chicago Press; Summa Publications; Univ. of Pennsylvania Press;  
Cornell Univ. Press; Cambridge Univ. Press; Univ. of Texas Press; Stanford  
Univ. Press; Duke Univ. Press; Univ. of Minnesota Press.

Journals

Studies in Romanticism; Tulsa Studies in Women's Literature; Philological  
Quarterly; Speculum; PMLA; International Journal of the Classical  
Tradition; Modern Language Quarterly; Journal of Medieval and Early Modern  
Studies

*International, National Offices and Committees:*

International Dante Seminar: Board of Directors, 1994-2003; Secretary, 1997-  
2003.

Modern Language Assoc.: Advisory Committee on the MLA International  
Bibliography, 1992-95 (Chair, 1993-95); Executive Committee, Division on Medieval and  
Renaissance Italian Literature, 1993-98; Executive Council, 1994-98; Committee on Academic  
Freedom, Professional Rights, and Professional Responsibilities, 2000-2003 (Chair, 2002-03)

Midwest Modern Lang. Assoc.: President, 1995.

Dante Society of America: Council, 1990-93; Nominating Committee (Chair), 1994-  
96.

Council of Colleges of Arts and Sciences: Committee on the National Agenda,  
1993.

*National Grant Review:*

National Endowment for the Humanities: Romance Languages Panel, Translation  
Division, 1982, 1984; Fellowships Division, 1994, 1995, 1996 (declined), 1997.

#### OTHER PROFESSIONAL ACTIVITIES:

University of Minnesota: Center for Medieval Studies, Organizer of Conference,  
"The Cultures of Papal Avignon," 2002.

North Central Association of Schools and Colleges: Consultant-Evaluator on  
Accreditation Review Teams, 1994-96; 1997-2001; 2001-2005.

Invited Lectures and Conference Papers: about 60 in all parts of North America,  
Italy, and France.

#### PUBLICATIONS (selected):

*The Comparative Perspective on Literature: Essays in Theory and Practice* (in collaboration with Clayton  
Koelb). Ithaca and London: Cornell University Press, 1988. viii + 378 pp. Far Eastern edition, Taipei:  
Bookman Books, 1988.

*Timely Reading: Between Exegesis and Interpretation*. Ithaca and London: Cornell University Press,  
1988. xv + 249 pp.

Tommaso Schifaldo, *Libellus de indagacionibus grammaticis*, in collaboration with Robert Kaster),  
*Humanistica Lovaniensia*, Vol. XXXII (1983), pp. 107-156.

"Dante e lo sviluppo delle istituzioni bancarie a Firenze: 'i subiti guadagni'" in Michelangelo Picone,  
ed., *Dante: Da Firenze all'aldilà*, Atti del terzo Seminario Dantesco Internazionale, Firenze, 2000,  
Firenze: Franco Cesati, 2001, pp. 249-261.

“Virility, Nobility, and Banking: The Crossing of Discourses in Dante’s *Tenzzone with Forese*,” in Teodolinda Barolini and H. Wayne Storey, eds., *Dante 2000: Themes and Problems*, New York: Fordham Univ. Press, 2004, pp. 241-258.

“Pétrarque polémiste anti-commercial?: ses portraits d’Avignon et de Venise,” Actes du colloque international, janvier 2004, Musée Pétrarque et Université d’Avignon et al., ed. Eve DuPerray, Paris: Beauchesne, 51-60.

“Medieval Texts and National Identities: Dante in Red, White, Green—Then Black,” *M/MLA Journal* 40:1 (Spring 2007), 11-24.

**Reviews** in *Italica*, *Studies in Romanticism*, *Gradiva*, *Journal of Religion*, *Speculum*, *American Journal of Semiotics*, *New Vico Studies*, *Journal of Medieval Latin*.

### **In preparation**

*Dante’s Divine Economies: Lineage, Wealth, and Prophecy in the “Paradiso.”*

Book ms.

*Petrarch’s Babylon: Cultural Intercourse in Papal Avignon*, Book ms. of essays collected and edited by S.N., with introduction by S.N.

### **WILLIAM G. THOMAS III**

University of Nebraska-Lincoln  
615 Oldfather Hall  
Department of History  
Lincoln, NE 68588  
[wgt@unl.edu](mailto:wgt@unl.edu)

### Employment

University of Nebraska-Lincoln  
John and Catherine Angle Chair in the Humanities and Professor of History  
University of Virginia  
Director, Virginia Center for Digital History, 1998-2005  
Associate Professor, Corcoran Department of History, 2004-2005  
Assistant Professor, Corcoran Department of History, 2001-2004  
Research Assistant Professor, 1997-2001  
Project Manager, Institute for Advanced Technology in the Humanities, 1996-1997

### Education

Ph.D.--University of Virginia, History, 1995  
M.A.--University of Virginia, History, 1991  
B.A.--Trinity College (Connecticut), History, 1986 with honors in History

### Selected Publications

"William Jennings Bryan, the Railroads, and the Politics of 'Workingmen'," *Nebraska Law Review* Vol. 86 No. 1 (2007): 161-179.  
"Black and on the Border," co-author with Edward L. Ayers and Anne Sarah Rubin, in Gabor Boritt, ed., *The African American Soldier in the Civil War* (Oxford: Oxford University Press, 2007)  
"The Countryside Transformed: The Eastern Shore of Virginia, the Pennsylvania Railroad, and the Making of a Modern Landscape," with Brooks Barnes and Tom Szuba, *Southern*

Spaces. July, 2007. < <http://www.southernspaces.org/contents/2007/thomas/1a.htm>>  
"Nothing Ought to Astonish Us: Confederate Civilians in the 1864 Shenandoah Valley Campaign," in Gary Gallagher ed. *The Shenandoah Valley Campaign of 1864* (Chapel Hill: University of North Carolina Press, 2006)  
"The Differences Slavery Made: A Close Analysis of Two American Communities," co-author with Edward L. Ayers, *American Historical Review*, December 2003.  
*The Civil War on the Web: A Guide to the Best Sites*, co-author with Alice E. Carter, *Scholarly Resources*, 2001.  
*Lawyering for the Railroad: Business, Law, and Power in the New South*, Louisiana State University Press, 1999.

#### Honors

OAH Distinguished Lecturer, 2007-present  
Mead Honored Faculty, University of Virginia, 2004-2005  
The James Harvey Robinson Prize, 2003, American Historical Association, with Edward L. Ayers and Anne S. Rubin, for *Valley of the Shadow: Two Communities in the American Civil War*  
The Lincoln Prize, 2001, Gettysburg College, Lincoln and Soldiers Institute, with Edward L. Ayers and Anne S. Rubin, for *Valley of the Shadow: Two Communities in the American Civil War*.  
Emmy Nominee, Washington, D. C. Chapter, National Television National Academy of Television Arts and Sciences, 2000, with George H. Gilliam, producer, for *The Ground Beneath Our Feet*, "Massive Resistance."

#### Board Service

Board of Editors, University of Nebraska Press, 2007-present  
Board of Editors, Southern Spaces, Emory University, 2003-present  
Institutional Review Board, University of Nebraska, 2006-present  
Research Advisory Board, University of Nebraska, 2006-present  
Board of Directors, Center for Liberal Arts, University of Virginia, 1999-2005

#### Recent Lectures and Conference Papers

"Fifty Years Later: The South, Virginia, and the Civil Rights Struggle," Alexandria Library Company Lecture in Southern History, March 2007.  
"Railroads, Civil War, and the Making of Modern America," American Historical Association, January 2007.  
"Time, Space, and History" with Edward L. Ayers, EDUCAUSE, Dallas, October 2006.

#### Electronic Archives and Publications (Director or Co-Director)

"Railroads and the Making of Modern America"  
<http://railroads.unl.edu>  
"Digital History"  
<http://digitalhistory.unl.edu>  
"The Valley of the Shadow: Two Communities in the American Civil War"  
<http://valley.vcdh.virginia.edu>  
"The Eastern Shore of Virginia and the Railroad, 1870-1930"  
<http://www.vcdh.virginia.edu/eshore>  
"Television News in the Civil Rights Era, 1950-1970"  
<http://www.vcdh.virginia.edu/civilrightstv>  
"The Geography of Slavery in Virginia"  
<http://www.vcdh.virginia.edu/gos>  
"Race and Place: An African American Community in the Jim Crow South, 1870-1920"

<http://www.vcdh.virginia.edu/afam/raceandplace/>

Current Grants and Fellowships

American Council of Learned Societies, Digital Innovation Fellowship, 2007-2008

Newberry Library, Short-term Fellowship, 2007-2008

Current Research Area

Jupiter's Bow: Railroads, The Civil War, and the Roots of Modern America--This book project (Yale University Press) examines the relationship between the railroad culture of the 1850s and 60s and the coming, fighting, and aftermath of the American Civil War. The book research will draw on and from the digital project on Railroads and the Making of Modern America (<http://railroads.unl.edu>).

**Appendix VII: Curriculum Vitae**  
**High Performance Computing Center Staffs**

**DIANE A. BAXTER**

**EDUCATION**

- 1975 BA - Biology, University of California, Los Angeles, CA  
1983 Ph.D. - Zoology (Botany minor), Duke University, Durham, NC, John Sutherland, Advisor.  
Dissertation Research: Population ecology of *Littorina irrorata* in a N.C. salt marsh

**APPOINTMENTS**

- 2/05 – Education Director, San Diego Supercomputer Center, UCSD, La Jolla, CA  
9/02-1/05 Development Director, Quail Botanical Gardens, Encinitas, CA  
9/00-9/02 Resource Development Specialist, Quail Botanical Gardens, Encinitas, CA  
4/99-9/00 Development Director, Pflieger Institute of Environmental Research, Oceanside, CA.  
8/98- 4/99 Education Partnerships Coordinator; Birch Aquarium at Scripps; Scripps Institution of Oceanography, University of California, San Diego  
2/84- 8/98 Curator of Education, Birch Aquarium, Scripps Institution of Oceanography (SIO), UCSD.  
1983 Coordinator, Public Understanding of Science Program, North Carolina Aquarium at Pine Knoll Shores, Atlantic Beach, N.C.  
1977- 83 Teaching Assistant, Duke University Zoology Dept. & Duke Marine Laboratory, N. C.  
1976 Research Associate, Botany Department, UCLA; Park S. Nobel, Supervisor

**RELEVANT PUBLICATIONS AND PRESENTATIONS**

- Baxter, D.; TeraGrid: Advancing Scientific Discovery and Learning; EDUCAUSE Annual Conference; October, 2007; Seattle, WA  
Baxter, D., Partnerships and Pathways to the *Internet to the Hogan and Diné Grid*; Society for the Advancement of Chicanos and Native Americans in Science (SACNAS) Annual Conference, October 2007, Kansas City, MO  
Mason, A. and Baxter, D., *Evolution of TeacherTECH*; TeraGrid 07; June, 2007; Madison, WI  
Baxter, D; *Power of Imagination: Next Generation Education*; Supercomputing 2006; Nov. 2006; Tampa, FL  
Baxter, D.; McGinnis, L; Wiziecki, E.; *Internships and Mentoring to Broaden Participation in Computing*; TeraGrid 2006; June, 2006; Indianapolis, IN  
Baxter, D.; *Education for a Changing World*; Supercomputing 2005; Nov. 2005; Seattle, WA  
Cacciola, J. and Snodgrass, R, 2002, *Sea Star On-Board Ocean Science Curriculum*, D. Baxter, Editor  
Education Department, Stephen Birch Aquarium-Museum, 1996, *Forecasting the Future: Exploring Evidence for Global Climate Change*, National Science Teachers Association Press, Arlington, VA

**SYNERGISTIC PROFESSIONAL ACTIVITIES**

- 2008 Education Program Co-Chair, TeraGrid '08 Annual Conference

2007 CI-TEAM Workshop Committee, July, 2007  
 2006-2007 SC-07 Education Program Committee  
 2006-2008 Board of Directors, Arroyo Paseo Charter School Foundation  
 2006 Volunteer Committee Chair, SC 06 Education Program  
 2006 Conference Committee, TeraGrid 2006 Conference, Indianapolis, IN  
 2005-2008 San Diego Science Leadership Council, San Diego County Office of Education  
 2005-2006 Advisory Board, Expanding Your Horizons, San Diego Science Alliance  
 1998-2006 Children's Garden Design Committee, Quail Botanical Gardens  
 1995-98 Board of Directors, San Diego Triton Project, U.S. Dept. of Education  
 1992-98 Board of Directors, San Diego Urban Systemic Initiative (NSF program)  
 1991-95 Steering Committee, San Diego's Project 2061, AAAS  
 1988-92 Board of Directors, National Marine Educators Association  
 1989 Marine Science Education Consortium, California Science Framework Committee

## COLLABORATIONS & OTHER AFFILIATIONS

### Current Projects:

*SCI: TeraGrid Resources Partners (HPCOPS);* Mark Sheddon, P.I.

*TeraGrid (GIG) - Education, Outreach, and Training*

Dane Skow, P.I., University of Chicago – Argonne National Laboratory; Scott Lathrop, Director, Education, Outreach, and Training; and External Relations

*CI-TEAM: Minority-Serving Institutions Cyberinfrastructure Empowerment Coalition*

Richard Alo, Principal Investigator; Geoffrey Fox, Co-PI; Alex Ramirez, Co-PI; Al Kuslikis, Co-PI; Selena Singletary, Co-PI; Diane Baxter, Co-PI

*IT-EESET: Information Technology – Engineering and Environmental Science Education Tools, 2007-2009*

Jeanne Ferrante, P.I.

*BPC-D Worlds for Information Technology and Science (SDSC Dissemination Site Lead)*

David Gries, P.I.; Margaret Corbit, Co-PI; Cornell University

*UC-Village Links, UCSD Faculty Research Project and Sixth College Practicum; 2007-2008;* Michael

Cole, P.I.; Laboratory of Comparative Human Cognition and Sixth College

### Prior Projects (past 48 months):

*Delivering Cyberinfrastructure: From Vision to Reality;* Fran Berman, P.I.

*CIP-EOT: CyberInfrastructure Partnership - Education, Outreach, and Training*

Fran Berman, P.I.; Thom Dunning, Co-PI;

*CI-TEAM: Minority-Serving Institutions Cyberinfrastructure Institute [MSI C(I)2]: Bringing Minority Serving Institution Faculty into the Cyberinfrastructure and e-Science Communities*

Geoffrey Fox gcf@indiana.edu (Principal Investigator); Richard Alo (Co-P.I.); Carrie Billy (Co-P.I.); Alexander Ramirez (Co-P.I.); Karl Barnes (Co-P.I.)

*EPIC; Expanding Participation in Computing*

Moses, Greg; University of Wisconsin, Madison (P.I.); Giles, Roscoe; Boston University (co-PI)

## SHAWN T. BROWN

Phone: (412) 268-4635

E-mail: stbrown@psc.edu

Web: <http://www.psc.edu>

## PROFESSIONAL PREPARATION

Bethany College Chemistry B.S. 1993-1997

University of Georgia Chemistry Ph.D. 1997-2001

## APPOINTMENTS

- Senior Scientific Specialist: Pittsburgh Supercomputing Center, Carnegie Mellon University, Mellon Institute (2005-resent)
- Research Associate: Q-Chem, Inc. Suite 690, 5001 Baum Blvd., Pittsburgh, PA 15213 (2001-2004)

## PUBLICATIONS (selected)

- “PSI3: An open-source ab initio electronic structure package.” Crawford T. D., Sherrill C. D., Valeev, E. F., Fermann J. T., King R. A., Leininger M. L., Brown S. T., Janssen C. L., Seidl E. T., Kenny J. P., Allen W. D. 2007, *J. Comp. Chem.*, 28, 1610.
- “Advances in methods and algorithms in a modern quantum chemistry program package”, Shao, Y, Molnar L. F., Jung, Y., Kussmann, J. Oschenfeld, C, Brown, S. T., Gilbert, A. T. B., et al 2006, *Physical Chemistry Chemical Physics B*, 8, 3172.
- “Enabling Computational Science on the Cray XT3”, Nystrom, N., Weisser, D., Lim, J., Wang, Y., Brown, S. T., Reddy, R., Stone, N.T.B., Woodward, P., Porter, D., Di Matteo, T., Kalé, L. V., Zheng, G.: Lugano 2006, CUG 2006 Proceedings Switzerland.
- “Efficient computation of the exchange-correlation contribution in density functional theory through multiresolution”, Kong, J, Brown, S. T., and Fusti-Molnar, L. 2006, *J. Chem. Phys.*, 124, 094109.
- “Interpolation of density values on a Cartesian grid: Improving the efficiency of Lebedev based numerical integration in Kohn-Sham density functional algorithms”, Brown, S. T. and Kong, J. 2006, *Chem. Phys. Lett.*, 408, 418.
- “IncDFT: Improving the Efficiency of Density Functional Theory Using Some Old Tricks”, Brown, S. T. and Kong, J. 2005, *Chem. Phys. Lett.*, 408, 395.
- “Unexpected encounters in simulations of the ALDH mechanism”, Hempel, J., Nicholas, H. B. Jr., Brown, S. T., Wymore, T. 2007, in “Enzymology and Molecular Biology of Carbonyl Metabolism”, 13, 9-13, Weiner, H., B. Plapp, R. Lindahl, and E. Maser, eds. Purdue University Press, Lafayette, IN.
- “A Combined Density Functional Theory and Molecular Mechanics (QM/MM) Study of FeCO

## SYNERGISTIC ACTIVITIES

- collaborates with Troy Wymore at National Resource for Biomedical Supercomputing (NRBSC) and Martin Field from the Institut de Biologie Structurale in Grenoble, France to enhance the QM/MM, parallel and grid-enabled capabilities of the DYNAMO molecular dynamics library.
- co-organizer and instructor for a workshop held by the NRBSC on QM/MM calculations held at the PSC in Sept, 2007.
- collaborates with Axel Kohlmeyer and Michael Klein from the University of Pennsylvania providing HPC instruction to students of their University.
- co-organized the 2nd annual QM/MM workshop held in August '07 at Univ. Penn
- actively involved in the PSC’s collaboration with the TeraGrid, playing a key role as Advanced Support Consultant.
- actively worked with research groups all over the world to bring research and computer codes to modern computer architectures in fields such as chemistry, biology, epidemiology, geology, and mathematics.
- involved in furthering the organization’s role in computer science through involvement in Extreme Scaling and HPC University working groups for the TeraGrid.

- serves as a scientific contact and consultant for the Computational Chemistry and Materials (CCM) functional area under the User Productivity Enhancement and Technology Transfer (PET) component of the DoD's High Performance Modernization Program (HPCMP). See <http://www.hpcmo.hpc.mil/Htdocs/PET/>.
- collaboration with researchers from Virginia Tech's Bioinformatics institute (VBI, Stephen Eubank and Madhav Marathe) and RTI International (Doug Roberts and Diglio Simoni) to provide develop a large scale epidemiological computation capability to the TeraGrid.

## COLLABORATORS

Wesley Allen, Univ. of Georgia

Jean Blaudeau, HPTi

Eric Bohm, UIUC

T. Daniel Crawford, Virginia Tech

Stephen Eubanks, Virginia Tech.

Martin Field, Institut de Biologie Structurale

Marek Freindorf, Univ. of NY at Buffalo

Thomas Furlani, Univ. of NY at Buffalo

Laszlo Fusti-Molnar, Univ. of Florida

Peter Gill, Australian National University

Mark Gordon, Iowa State University

Martin Head-Gordon, Univ. of Calif.,

John Hemple, Univ. of Pittsburgh

Curtis Jenssen, Sandia National Labs

L. V. Kale, UIUC

Nicholas Karonis, N. Illinois University

Gary Kedzora, HPTi

Joseph P. Kenny, Sandia National Laboratory

Jing Kong, Q-Chem, Inc.

Anna Krylov, Univ. of Southern California

Jerzy Leszczynski, Jackson State University

Madhav Marathe, Virginia Tech.

Shirley Moore, University of Tennessee, Knoxville

Hugh Nicholas, PSC

Nicholas Nystrom, PSC

Andrew Rappe, University of Pennsylvania

Doug Roberts, RTI

Tony Rollett, Carnegie Mellon University

Yihan Shao, Q-Chem, Inc.

Berkeley David Sherrill, Georgia Tech.

Diglio Simoni, RTI

Joel Stiles, PSC

Edward Valeev, Virginia Tech.

Deborah Weisser, Cray, Inc.

Troy Wymore, PSC

Anthony Yau, HPTi

## GRADUATE AND POST DOCTORAL ADVISORS

Henry F. Schaefer III, graduate advisor, University of Georgia, Athens, GA

## LAURA F. MCGINNIS

Pittsburgh Supercomputing Center

Carnegie Mellon University

300 South Craig St, #313

Pittsburgh, PA 15213

412-268-5642

LFM@psc.edu

### a. Professional Preparation

University of Pittsburgh Computer Science B.Sc., 1982

Carnegie Mellon University Industrial Administration M.Sc., 1999

## **b. Appointments**

2006-present Project Manager, Systems & Operations, PSC  
1999-2006 Project Coordinator, Data & Information Resource Services, PSC  
1996-1999 Senior Database Administrator, Management Information Systems, PSC  
1993-1996 Senior Software Engineer, Computing Services, Carnegie Mellon University  
1990-1993 Senior Software Engineer, Administrative Computing and Information Systems, Carnegie Mellon University  
1989 Consultant, Computer Specialists, Inc., Pittsburgh PA  
1987-1989 Manager, Technical Services: Strategic Financial Group  
1984-1986 Programmer/Analyst, Support Consultant, Telesis Computer Corporation, Pittsburgh, PA  
1982-1984 Database Programmer/Analyst, University of Pittsburgh Medical Center, Pittsburgh, PA  
1982-1984 Part-time Instructor, Department of Computer Science, University of Pittsburgh

## **c. Publications**

1. R. Mach, R. Lepro-Metz, S. Jackson, L. McGinnis . “Usage Record – Format Recommendation”; Open Grid Forum Document Series, GFD.98 (approved March 2007).
2. Catlett, Charlie, and 95 others. “TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications,” in *High Performance Computing and Grids in Action*, ed. L. Grandinetti, IOS Press of Amsterdam (to appear).

## **d. Synergistic Activities**

TeraGrid

- HPC University Requirements Analysis Team – Chairperson
- HPC University Working Group – Member
- Education, Outreach, and Training Working Group – Training Coordinator
- TG06, TG07 National Conferences – Birds-of-a-Feather Chairperson, committee member
- TG08 National Conference – Poster Session Chairperson
- Core Service 2.0 Project – Component Coordinator Open Grid Forum
- Usage Record Working Group: Chairperson; editor of the Usage Record Schema specification, v.1.
- Production Grid Services Research Group: Co-Chairperson Supercomputing Science Consortium
- Grids and Clusters Working Group chairperson

## **e. Collaborators and Other Affiliations**

*Collaborators and Co-Editors*

Catlett, Charlie Argonne National Laboratories  
Hart, Dave San Diego Supercomputer Center  
Milfeld, Kent Texas Advanced Computing Center  
Quinn, Steve National Center for Supercomputing Applications  
Skow, Dane University of Chicago/Argonne National Lab  
Towns, John National Center for Supercomputing Applications

## **JOHN URBANIC**

### (a) Professional Preparation

Carnegie Mellon University Physics B.S. 1989  
Pennsylvania State University Physics M.S. 1991

### (b) Appointments

1998-2007 Staff Computational Science Consultant, Pittsburgh Supercomputing Center  
1993-1998 Sr. Computation Science Consultant, Pittsburgh Supercomputing Center

1991-1993 Computational Science Consultant, Pittsburgh Supercomputing Center

(c) Publications (selected)

Kemal Ebcioglu, Vivek Sarkar, Tarek El-Ghazawi, John Urbanic. An Experiment in Measuring the Productivity of Three Parallel Programming Languages. Proceedings of the Third Workshop on Productivity and Performance in High-End Computing (PPHEC-06) Held in conjunction with the Twelfth International Symposium on High Performance Computer Architecture, 2006.

Nystrom, N.A., Urbanic, J., and Savinell, C. Understanding Productivity Through Nonintrusive Instrumentation and Statistical Learning. Proceedings of the Second Workshop on Productivity and Performance in High-End Computing (PPHEC-05), 2005. 53-61.

Jacobo Bielak, Omar Ghattas, Julio C. Lopez, Kwan-Liu Ma, David R. O'Hallaron, Leonardo Ramirez-Guzman, Nathan Stone, Ricardo Taborda-Rios, John Urbanic. Analytics challenge---Remote runtime steering of integrated terascale simulation and visualization, Proceedings of the 2006 ACM/IEEE conference on Supercomputing, Tampa, Florida, 2006 ISBN:0-7695-2700-0

Nick Nystrom, Deborah Weisser, and John Urbanic. The SUMS Methodology for Understanding Productivity: Validation Through a Case Study Applying X10, UPC, and MPI to SSCA#1. Proceedings of the Third Workshop on Productivity and Performance in High-End Computing (PPHEC-06), 2006.

D. C. O'Neal and J. Urbanic, On Microprocessors, memory hierarchies, and Amdahl's law, Proceedings of the DoD HPCMP Users Group Conference, Monterey, CA, 1999.

Volkan Akcelik, Jacobo Bielak, George Biros, Ioannis Epanomeritakis, Antonio Fernandez, Omar Ghattas, Eui Joong Kim, Julio Lopez, David O'Hallaron, Tiankai Tu, and John Urbanic. High Resolution Forward and Inverse Earthquake Modeling on Terascale Computers, Proceedings of ACM/IEEE SC2003, Phoenix, AZ, 2003 ISBN:1- 58113-695-1

Nikhil Kelshikar, Xenophon Zabulis, Jane Mulligan, Kostas Daniilidis, Vivek Sawant, Sudipta Sinha, Travis Sparks, Scott Larsen, Herman Towles, Ketan Mayer-Patel, Henry Fuchs, John Urbanic, Kathy Benninger, Raghurama Reddy, , and Gwendolyn Huntoon. Real-time Terascale Implementation of Tele-immersion. In Proc. of the Terascale Performance Analysis Workshop, Melbourne, Australia, June 2003. In conjunction with ICCS'2003 - International Conference on Computational Science (LNCS 2660)

(d) Synergistic Activities

Teach multiple workshops and seminars on massively parallel computing each year. This training specifically targets scientists, not programmers, and very large platforms.

Developed with Carnegie Mellon Robotics Institute Eyevision™ system used by CBS in 2001 Superbowl and succeeding broadcasts. Responsible for real-time camera robot control.

(e) Collaborators & Other Affiliations (2003-2007)

- Collaborators and Co-Editors.
  - o Volkan Akcelik Carnegie Mellon University
  - o Kathy Benninger Pittsburgh Supercomputing Center
  - o Jacobo Bielak Carnegie Mellon University
  - o George Biros University of Pennsylvania
  - o Kostas Daniilidis University of Pennsylvania
  - o Kemal Ebcioglu International Business Machines
  - o Ioannis Epanomeritakis Carnegie Mellon University
  - o Antonio Fernandez Carnegie Mellon University

- o Henry Fuchs UNC Chapel Hill
- o Omar Ghattas University of Texas, Austin
- o Tarek El-Ghazaw George Washington University
- o Gwendolyn Huntoon Pittsburgh Supercomputing Center
- o Nikhil Kelshikar University of Pennsylvania
- o Eui Joong Kim Carnegie Mellon University
- o Scott Larsen UNC Chapel Hill
- o Julio C. Lopez Carnegie Mellon University
- o Kwan-Liu Ma University of California, Davis
- o Ketan Mayer-Patel UNC Chapel Hill
- o Jane Mulligan University of Colorado at Boulder
- o Nystrom, N.A Pittsburgh Supercomputing Center
- o David R. O'Hallaron Carnegie Mellon University
- o D. C. O'Neal Pittsburgh Supercomputing Center
- o Leo Ramirez-Guzman Carnegie Mellon University
- o Raghurama Reddy Pittsburgh Supercomputing Center
- o Vivek Sarkar International Business Machines
- o Vivek Sawant University of Pennsylvania
- o Savinell, C. Pittsburgh Supercomputing Center
- o Sudipta Sinha UNC Chapel Hill
- o Travis Sparks UNC Chapel Hill
- o Nathan Stone Pittsburgh Supercomputing Center
- o Ricardo Taborda-Rios Carnegie Mellon University
- o Herman Towles Herman Towles
- o Tiankai Tu Carnegie Mellon University
- o Deborah Weisser Pittsburgh Supercomputing Center
- o Xenophon Zabulis University of Pennsylvania

**Appendix VIII: Previous Workshop Experiences**

Date	Workshop	Location	Participants	Sponsoring Organization
Dec-05	Introduction to Digital Humanities	NCSA	100	I-CHASS
February 2006 (repeated 2007 2008)	Text Encoding Initiative	NCSA	20 per year	I-CHASS and GSLIS
Mar-06	Using iLife 06 in Teaching and Research	NCSA	50	Apple, Inc.
Summer 2006	Cyberinfrastructure in the Humanities, Arts, and Social Sciences	SDSC	60	National Science Foundation/EPIC
Sep-06	Katrina: After the Storm – Civic Engagement Through Arts, Humanities and Technology	NCSA	150	I-CHASS/NCSA/University of Illinois
Dec-06	Spatial Thinking in the Social Sciences and Humanities	NCSA	80	I-CHASS/EPIC
Jun-07	e-Science for Arts and Humanities Research: an Early Adopters Forum	NCSA	25	I-CHASS/King’s College, London
Jun-07	Digital Humanities07	NCSA	400	GSLIS /I-CHASS/NCSA/University of Illinois
Summer 2007	Computational Methods in Humanities, Arts, and Social Science	SDSC	35	Supercomputing 07 Education Committee/TeraGrid
Apr-08	SEASR Mini-Residencies	NCSA	25	Mellon Foundation/SEASR/NCSA/Illinois Informatics Initiative (I3)
Apr-08	Supercomputing 08 Planning Workshop	NCSA	20	Supercomputing 08 Education Committee
July 13- July 19, 2008	Bio-Nano-Info-Socio Workshop	San Jose, Costa Rica	~30	I-CHASS, NCSA, ARTCA
July 27- August 3, 2008	High Performance Computing	NCSA	~35	Supercomputing 08 Education Committee
Sep-08	Data-Mining in the Humanities	NCSA	~45	Mellon Foundation/SEASR/NCSA/Illinois Informatics Initiative (I3)

## **Appendix IX: Computational and Logistical Resources**

### **Facilities, Equipment and Other Resources**

NCSA continues to support user communities by offering the resources that are the foundations of advanced cyberinfrastructure. The total computational resources exceed 145 TF supported by over 1.3 PB of disk storage as part of the infrastructure. The systems are on an internal multi-10GigE network. Below is a summary of those resources.

#### **NCSA Compute Resources**

##### **Abe**

Clovertown 2.38Ghz Blades from Dell  
Interconnect: InfiniBand  
1200 blades, 2400 processors (quad core)  
8GB of memory per blade (1 GB/core), 9.6TB total  
Peak performance: 89.5TF  
170 TB Lustre filesystem

##### **Mercury, Phase 1 (Hg 1)**

Itanium 2 1.3 GHz IBM Linux cluster  
Interconnect: Myrinet 2000  
256 nodes, 512 processors  
4 GB and 12 GB memory/node, 2.0 TB total  
Peak performance: 2.6 TF  
100 TB GPFS filesystem, NFS, Lustre WAN, GPFS WAN

##### **Mercury, Phase 2 (Hg 2)**

Itanium 2 1.5 GHz IBM Linux cluster  
Interconnect: Myrinet 2000  
667 nodes, 1334 processors  
4 GB memory/node, 2.5 TB total  
Peak performance: 8 TF  
100 TB GPFS filesystem, NFS, Lustre WAN, GPFS WAN

##### **Cobalt (Co)**

SGI Altix systems, 2x512 processors Itanium 2 1.6 GHz systems, Linux  
Interconnect: SGI, Numalink  
1,024 processors  
4 TB total memory  
Peak performance: 6.6 TF  
250 TB SAN storage with SGI CxFS filesystem  
8 x 8p SGI Prism visualization systems with Infiniband interconnects to the 512p SMPs

##### **Tungsten**

Intel Xeon 3.2 GHz Dell Linux cluster  
Interconnect: Myrinet 2000  
1280 nodes, 2560 processors  
3 GB memory/node, 3.8 TB total  
Peak performance: 16.4 TF

122 TB Lustre filesystem

### **T3**

Woodcrest 2.66 Ghz Blades from Dell

Interconnect: InfiniBand

520 Blades, 1,040 processors (dual core)

8GB of memory per blade (2 GB/core), 4.1 TB total

Peak performance: 22.1 TF

20 TB Lustre filesystem

Primarily used by NCSA Private Sector Program Partners

### **Mass Storage**

The environment currently consists of 3 SGI Origin 3900 servers running EMC/Legato DiskExtender (UniTree) with 180TB of SAN disk cache, 38 LTO2 tape drives, 14 IBM LTO3 tape drives, and 1 ADIC library. The total archival storage capacity of this environment is 5 PB.

### **Infrastructure SAN**

384 TB of SAN connected storage for infrastructure and special projects. This utilizes a high availability SAN configuration allowing for multiple paths to the storage depending on applications needs to access data. Backups and other configuration parameters can be added depending on applications needs.

### **High Performance Network**

All computing platforms are interconnected to a multi-10gigabit network core. NCSA's highperformance computing environment has access to Abilene via a 10-gigabit-per-second connection. NCSA also is one of the leading sites for I-WIRE, an optical networking project funded by the state of Illinois. I-WIRE provides lambda services for several projects, including NCSA's 30-gigabit-per-second connection to the TeraGrid network.

### **Display Systems**

*Tiled Display Wall:* This environment consists of 40 NEC VT540 projectors, arranged in a matrix 5 high and 8 across. The output of the NEC VT540s is rear-projected towards a single screen, creating a large-format, high-resolution image space that is 8192 x 3840 pixels. A 40-node PC Linux cluster is used to drive the display wall. The machines are dual-processor Intel Xeons, running at 2.4 GHz, with Nvidia FX 5800 Ultra graphics accelerator cards, and communicating over Myrinet.

*High Definition Passive Stereo Theater:* The NCSA High Definition Passive Stereo Theater is a 1920x1080 display on an 6' x 3' x 5' screen. The projectors used are JVCD-1LA. The display is driven by a dual AMD Opteron 242 processor running at 1.6 GHz. Graphics hardware consists of a Nvidia Quadro FX3000.

### **Applications Software**

NCSA offers a variety of third-party applications and community codes that are installed on the High-performance systems at NCSA. These applications cover a wide range of science and engineering domains, data analytics and visualization, mathematics and statistics. Complete information on the packages available and detailed descriptions of them are available at:

<http://hpcsoftware.ncsa.uiuc.edu/Software/user/index.php?view=NCSA>.

**Conferences & Institutes**

A division of the University of Illinois' Office of Continuing Education , Conferences & Institutes will offer program development assistance; coordination of logistics; coordination of registration services, including the creation and management of an online registration database and on-site management; coordination of all marketing and promotional materials and services; establishment and management of a University account to collect registration income and pay conference expenses; and an itemized financial report after the institute.

## **Appendix X: Letters of Commitment and Support**

Note: The Letters of Support were signed and collected by I-CHASS prior to the withdrawal of the Texas Advanced Computing Center (TACC) from this H<sup>P</sup>C grant. A research center at the University of Texas at Austin, TACC provides advanced computing resources and services to enable computationally-intensive research via the development, operation, and support of advanced computing technologies. TACC's curriculum outlined a program of study in its "scientific visualization" technology, EnVision. EnVision dramatically simplifies the process of turning large amounts of numerical data into an intuitive, immersive, and interactive visual display. In its notification of withdrawal, TACC determined that they were unable to participate at this time because the technology had not been adequately adapted to humanities research. Group Leader Brad Armoskey related that TACC would like I-CHASS to consult on the adaptation of this resource for humanities users and would join in the next stage of this grant. The Support offered by the enclosed individuals remains consistent despite this change.

---

**Citation:**

Guiliano, Jennifer, and Simon Appleford. "Humanities High Performance Computing Collaboratory " National Endowment for the Humanities, Grant Submission, University of South Carolina, Columbia, SC, 2010.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of South Carolina  
Center for Digital Humanities  
University of Illinois  
Institute for Computing in Humanities, Arts, and Social Science

**Team members:**

University of South Carolina  
Jennifer Guiliano  
Jun Zhou  
Jijun Chang  
J.J. Shepherd  
University of Illinois  
Simon Appleford  
Alan Craig  
Alex Yahja  
Kenton Henry

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

**Significance:**

*I think we suffer from being a field in which an old boys (and a few women) network formed because there are few formal ways that people can train.<sup>1</sup>*

- Geoffrey Rockwell, June 9, 2010

Over the past twenty years, the interdisciplinary field of humanities computing has created a riot of research information and tools, developed by many institutions across a variety of technical formats and platforms. As Geoffrey Rockwell, Professor of Philosophy and Humanities Computing at the University of Alberta, Canada, has pointed out, however, there remain few formal programs for digital humanities training. The proposed **Humanities High Performance Computing Collaboratory (H<sup>P</sup>C)** project will address this significant gap by offering a collaborative education structure that serves as a resource for humanities scholars to receive theoretical and technical education in the digital humanities. Bridging the divide between digital humanities centers, supercomputing centers, and the innovative work of humanities scholars, the **Center for Digital Humanities (CDH)** at the University of South Carolina will lead a collaboration partnering the **Institute for Computing in Humanities, Arts, and Social Science (I-CHASS)** at the University of Illinois at Urbana-Champaign and the **National Center for Supercomputing Applications (NCSA)** that fosters an educational curriculum for non-specialists to enter into digital humanities research. H<sup>P</sup>C will engage humanists in dialogue with colleagues at CDH, I-CHASS, and NCSA, in order to enhance and accelerate humanities research and education for those with little to no digital humanities expertise and few local resources. Unlike most training programs that are run via institutes or workshops, H<sup>P</sup>C will not only provide educational training, it will also deploy sustained technical resources for scholars from low-infrastructure institutions in the form of portal design and hosting, computational storage, and technical staff. While some institutions are fortunate to have local computational expertise and robust technical/infrastructure that can be deployed for digital humanities research, many humanities faculty who are interested in digital humanities are daunted by their lack of institutional resources. H<sup>P</sup>C will bring these faculty into the fold by aligning their interests with consortium partners so that each participant will: 1) receive education in five computational concentrations; 2) receive instruction in digital humanities project design and management; 3) obtain hands on experience with a variety of technical platforms; 4) work with technical staff to outline pilot explorations in at least one area of computational concentration; and 5) join a year-long virtual community where scholars will support their peers in authoring a digital humanities project. Significantly, we will culminate our training by asking each participant to establish short- and long-term technological goals for each project that will “scale” the project’s development. CDH will serve as the coordination point for all H<sup>P</sup>C participants. This grant will facilitate two five-day residencies and a culminating two-day conference for thirty-five humanities participants as well as ten computing specialists at a cost of \$249,588. Twenty-five nationally selected participants will attend both training workshops and the culminating closing workshop, while five additional participants per workshop will be drawn from local scholars who will attend only their local workshop and the closing workshop.

This high performance computing-digital humanities partnership will integrate hardware for computing, data and networks, and an interoperable suite of software and middleware services and tools available at NCSA, I-CHASS, and CDH with the specialized needs of humanities researchers. We initiate this partnership as a result of our considerations of how humanities research is structured: throughout the lifetime of a digital research project the technical needs of the scholars evolve. While some scholars have ready access to digital humanities centers and supercomputing centers, there is an entire untapped population of humanists who do not have local training opportunities and computational infrastructure. Further, mobile applications, serious gaming, augmented reality, large-scale imaging, and social networking have emerged as fields of importance for humanities scholars looking to innovate new research questions and these are technical competencies that may not be readily available locally. Effective educational developments undertaken in this grant will allow humanities scholars to focus their intellectual and scholarly energies on acquiring a broad understanding of technological trends in digital

humanities research while simultaneously training them to be effective builders of new digital humanities projects. Significantly, each partner within this grant offers its own strength: NCSA has the ability to offer high-end technical resources that are beyond the reach of most researchers; I-CHASS offers experience in large-scale image analytics, social networking, and augmented reality as well as “lessons learned” from the previous iteration of this grant, awarded in 2008<sup>2</sup> (Appendix: Previous IATDH Workshop Experiences); CDH contributes its expertise in mobile application and serious gaming, two areas recognized by the 2011 Horizons Report as emerging fields for teaching, learning and creative inquiry<sup>3</sup>, as well as the experience of Co-Principal Investigators Dr. Jennifer Guiliano, Associate Director of CDH and Research Assistant Professor of History, and Dr. Jijun Tang, Associate Professor of Computer Science and Engineering. Dr. Guiliano has spent the last two years authoring formal and informal training for humanities scholars interested in digital humanities and high performance computing while Dr. Tang has led efforts at the University of South Carolina to integrate high performance computing education with humanities gaming and mobile platform education.

In considering this iteration of the H<sup>P</sup>C Consortium, we highlight the 2008 commitment of H<sup>P</sup>C: the power of high performance computing and digital humanities lies in offering innovative education and training in emergent technologies that responds to evaluation and assessment of previous iterations as well as consultations among the constituent centers about new horizons for digital humanities and high performance computing. Dr. Kevin Franklin, Primary Investigator<sup>2</sup> of the 2008 HPC grant and Executive Director of I-CHASS, worked with Dr. Guiliano in constructing the Collaboratory. While first-cycle participants were pleased with the computational education they received, the structure of the training program was prohibitive to sustained multi-site attendance. With few exceptions, each research group would substitute members in and out of the educational training as few scholars could afford to miss four weeks of teaching or research activities. There were few attendees who were able to attend all three of their mini-residences offered by HPC as well as the closing conference; thus most participants received only a portion of the slated educational curriculum. As a result, while individual participants evaluated the training and offered praise for their knowledge growth, the Primary Investigators were dissatisfied with being unable to offer all individuals a complete snapshot of digital humanities and high performance computing that led to clearly defined strategies for short and long-term project growth. Unsurprisingly, the most successful participants in the eyes of the organizers in creating these development trajectories were those who were attendees at all three residencies. Offering two intensive five-day workshops, the dates of which will be announced as part of the public call for participants, HPC will address this issue by requiring attendance at both educational workshops and the closing two-day workshop. In effect, we will reverse the paradigm from our first iteration where we selected promising digital humanities participants first and then attempted to schedule the workshops around their availability, to instead arrange the workshop schedule first and then solicit participants who can guarantee their attendance.

H<sup>P</sup>C will consist of a core set of presentations, hands-on sessions, and readings in computational technologies for humanities activities. Invited scholars will experience a day of training in a particular sub-field (Imaging Analytics, Social Networking, Mobile Applications, Augmented Reality, and Serious Gaming) with the fifth-day including a holistic integration of technologies with strategies for project design and management. By the close of H<sup>P</sup>C, each scholar will have a clearly delineated pathway for the inclusion of computing technologies into their work and will have been given the skills to successfully author and manage their own digital humanities and high performance computing projects. Our culminating workshop will bring together all participants and staff to present their projects, receive individual consultation with technical staff, and participate in discussions regarding the future of digital humanities and high performance computing within the academy. We see an Institute of this type as essential to the future of digital humanities as: 1) there are significant portions of the humanities constituency that are unengaged in digital humanities and high performance computing; 2) while there are some humanists exploring digital humanities and high performance computing, few receive formal training across these areas; 3) regardless of their expertise, all humanities scholars should be educated in

project design, management, and integration with an attention to digital humanities and supercomputer center partnerships; 4) cross-center collaboration needs to become an essential feature of digital humanities work just as it has utilized in the sciences.<sup>4</sup>

#### Institutional Profiles:

The **Center for Digital Humanities (CDH)** at the University of South Carolina was founded in 2008 by a group of scholars committed to the belief that the future of research and teaching in the humanities is bound to digital methodologies, pedagogies, and media. Drawing on its partnerships with faculty in the College of Arts & Sciences, the School of Library and Information Science, the University Libraries, the Department of Computer Science, and the South Carolina Digital Library, CDH initiates, nurtures, and assists in the development and deployment of computational resources, collaborative tools, and educational programming. Our pilot effort, the Digital Humanities Initiative (2008-2010), demonstrated the wide-ranging curiosity and spirit of innovation that form the core of this mission. We promoted individual projects featuring diverse technical challenges, fostered innovative uses of technology in the classroom, and served as an incubator for campus-wide efforts to promote digital technologies. Formally recognized in June of 2010 by the University of South Carolina Board of Trustees, CDH offers expertise in mobile applications, humanities gaming, imaging analytics, and social media.

*Organizational Structure:* CDH offers a core staff of four employees who work with humanities faculty, computational specialists, and independent scholars from across the Southeast. The Director of CDH, Dr. David Lee Miller, and the Associate Director, Dr. Jennifer Guiliano report to the University of South Carolina Dean of Arts & Sciences, Dr. Mary Anne Fitzpatrick, as well as a Campus Advisory Board. The Associate Director ensures the day-to-day running of the Centers's operations and consults regularly with the Director, the Campus Advisory Board, and USC leaders to discuss issues surrounding the Centers's strategic direction. The CDH staff (Appendix V: Curriculum Vitae- CDH Faculty and Staff) is complemented by the considerable experience and expertise of 75 plus researchers, technology specialists, and staff who work at USC and from whom CDH is able to draw upon as opportunities arise.

*Facilities and Resources:* CDH is housed in the College of Arts & Sciences and its Research I institution, the University of South Carolina. As such, it is able to utilize existing Center resources (concentrated in mobile applications, humanities gaming, and image analytics) as well as tap into the advanced resources afforded by its campus relationships. Founded in 2008 as the only dedicated digital humanities center in South Carolina, CDH is uniquely positioned to strengthen digital humanities scholarship for scholars who need a complete curriculum in digital humanities and high performance computing through its partnerships with I-CHASS and NCSA. The computing expertise and resources available at NCSA, harnessed and channeled for the humanities through CDH and I-CHASS, can create comprehensive interactive education and study in the humanities that will lead to insights into integrating individual researchers and their home institutions more fully into supercomputing.

The **University of Illinois at Urbana-Champaign** has long been at the forefront of science, technology, and burgeoning computer power beginning 15 years ago when MOSAIC, the world's first graphical web browser, was developed at the National Center for Supercomputing Applications (NCSA). MOSAIC has revolutionized research and teaching because it democratized access to information. Now bringing the same advances to the humanities, I-CHASS charts new ground in high-performance computing and the humanities, arts, and social sciences by creating both learning environments and spaces for digital discovery. Founded in 2004, I-CHASS presents path-breaking research, computational resources, collaborative tools, and educational programming to showcase the future of the humanities, arts, and social sciences by engaging visionary scholars from across the globe to demonstrate approaches that interface advanced interdisciplinary research with high-performance computing. I-CHASS provides these researchers with leadership-class computational resources, both human and technical, to enhance their knowledge discovery and exploration. I-CHASS maintains strategic partnerships with the NCSA, the Great Lakes Consortium for Petascale Computation (GLCPC)<sup>5</sup>, and the Illinois Informatics Institute (I<sup>3</sup>)<sup>6</sup>. These alliances uniquely position I-CHASS to offer unprecedented intellectual and technical

expertise to aid humanities researchers in producing interdisciplinary research solutions that will positively affect the human condition and experience. Further, as partners in TeraGrid, an open scientific discovery infrastructure combining leadership class resources at eleven partner sites to create an integrated, persistent computational resource, NCSA has resources that include more than 750 teraflops of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Researchers can also access more than 100 discipline-specific databases.

*Organizational Structure:* I-CHASS currently maintains a core staff of seven employees who work closely with humanities faculty, computing researchers, and research groups from across the globe. The Executive Director, Dr. Kevin Franklin, leads research and operations at I-CHASS. I-CHASS is guided by an Advisory Board of NCSA, UIUC campus and external leaders who discuss issues surrounding the Institute's strategic direction. The I-CHASS staff is complemented by the considerable experience and expertise of the 250-plus researchers, technology specialists, and staff who work at NCSA and from whom I-CHASS is able to draw as opportunities arise.

*Facilities and Resources:* Founded in 2004 as the only Digital Humanities Center birthed and hosted by a national supercomputer center, I-CHASS is uniquely positioned to strengthen digital humanities scholarship through its expansive technical networks. I-CHASS is housed at NCSA and its Research I institution, the University of Illinois. NCSA is invaluable by bringing massive computational resources (Appendix II: Computational and Logistical Resources) to the humanities by enabling new and more complex projects nationwide. NCSA employs top-level experts in every field crucial to humanities computing: sophisticated search and retrieval, data management and visualization, human-computer interaction, distributed, collaborative computing, and large-scale modeling and simulation. The computing expertise and resources available at NCSA, harnessed and channeled for the humanities through I-CHASS, can create cyberenvironments for digital-humanities research and entire virtual worlds for interactive education and study in the humanities. Further, as partners in TeraGrid, an open scientific discovery infrastructure combining leadership-class resources at 11 partner sites to create an integrated, persistent computational resource, NCSA and TACC have resources that include more than 750 teraflops of computing capability and more than 30 petabytes of online and archival data storage, with rapid access and retrieval over high-performance networks. Researchers can also access more than 100 discipline-specific databases.

The alliance between CDH, NCSA, and I-CHASS uniquely positions this collaborative to offer unprecedented intellectual and technical expertise to aid humanities researchers in producing interdisciplinary research solutions that will positively affect the human condition and experience.

### **Curriculum Overview:**

There will be two five-day long residencies held by technical staffs at I-CHASS/NCSA and CDH. CDH will host their five-day workshop with a curriculum concentrated on serious gaming, mobile applications, and project design/management while NCSA will host their five-day workshop on social networking and media, augmented reality, and large-scale image analytics. Mobile applications are an emergent field in digital humanities due to its ubiquity and portability. As "market saturation" for mobile phones as ownership rates amongst late teens approach 100%, there is a corresponding investment in utilizing mobile applications as both a mechanism for disseminating digital humanities projects and as a tool for digital humanities research. Curriculum for mobile application development will include learning of platforms for data management, hands on experience with Android and iPhone including exploring programming principles, and the issues of information deployment on mobile devices. Our humanities gaming curriculum will outline principles of gaming construction, hands on experience with existing gaming platforms including Flash, Unity, and Torque, and theoretical exercises in game construction. I-CHASS/NCSA will provide two days each of training in image analytics and augmented reality with their fifth day focusing on social networking and new media. Researchers in the social sciences and humanities are increasingly using computers to manage, organize and analyze non-numerical data from

textual sources including images, manuscripts, and video. The I-CHASS/NCSA components would examine technologies for image analyses and augmented reality. Computer technologies available to humanities scholars would include: (a) integration of historical spatio-temporal data with maps and web-based interfaces (georeferencing, spatial and temporal sampling, sub-setting, tiling and stitching, web-based open layers and server); (b) automated analysis of scans of historical manuscripts (color spaces, image statistics, classification, cropping); (c) 3D imaging using high resolution 2D images of historical artifacts or 2D videos (3D imaging principles, stereopsis, calibration, spectral properties); (d) analyses of large volumes of contemporary PDF documents (PDF document structure, information extraction and cleansing, clustering, versioning); and (e) self-describing executions of analyses using advanced workflow studio (preservation, scripting & workflow, provenance, tagging, distributed data & tools & computers.)

Participants will be required to attend both five-day workshops and the closing two-day event in order to receive a complete educational curriculum. An itemized curriculum is outlined in Appendix II: Curriculum Workplan Itinerary.

### **Participants:**

CDH will solicit participants via a national call that, while tapping existing digital humanities networks like HASTAC and centerNet for dissemination, will also target traditional humanities networks including H-Net, the major professional humanities organizations, and state-based humanities alliances. Significantly, we also intend to send directed invitations to submit an application to faculty at Historically Black Colleges and Universities and Tribal Colleges. Participants will be asked to provide a one-page description of their disciplinary interests, their experience with digital humanities and high performance computing, and an assessment of the local resources available to them. Applications will be blind reviewed by a four-person team consisting of University of South Carolina Co-PI Dr. Jijun Tang, University of Illinois curricular leader, Dr. M. Scott Poole, and two senior humanities scholars drawn from the CDH and I-CHASS Campus Advisory Boards. Primary Investigator Dr. Guiliano will gather all applications and ensure the fairness of the review process. As we are hoping to draw a wide range of participants from across a broad array of academic backgrounds and experience with digital humanities, we do not intend to impose a predetermined set of selection criteria, preferring instead to allow applicants to make the case for their inclusion in the workshops themselves.

### **Impact and Evaluation:**

Findings from this project will be disseminated in traditional and innovative ways. We will encourage and maintain collaboration among humanities and high performance computing participants by offering a web-portal that allows electronic dissemination and maintains a constant web-based presence. The primary goal of disseminating experience, breakthroughs and learned lessons will be achieved by putting the curriculum and associated publications online. The secondary goal will be facilitated by the inclusion of modules embedded in the portal that allows activities tracking (provenance data), collaboration support (including blogs, chat, and wikis), and networking support. Creating a multi-way networked activity centered on the digital humanities, the experience, breakthroughs, and lessons learned from each project will be disseminated to the wider humanities academic audience and the general public through our virtual community in addition to the more traditional online papers, journal articles and research reports. Our ambition is to achieve continual and energetic discussion and collaboration for each participant and as a collaborative. HPC involves a unique combination of assets: 1) it will further the educational efforts of the selected humanities scholars; 2) it will allow for the refinement of technology education by computational scholars; 3) it will create a collaborative infrastructure and virtual community accessible to scholars across humanities. The participants will author their own project planning memorandum as well as offer recommendations for the next iteration of the Consortium. In serving not just invited participants but also scholars interested in the humanities and digital technologies, HpC captures the underlying intent

of the National Endowment for the Humanities grant programs: to interest people in the Humanities and aid them in their quest to more fully understand human life and experiences.

Within the confines of HpC, our evaluation process will be conducted via evaluation tools provided by the Center for Digital Humanities and the Institute for Computing in Humanities, Arts, and Social Sciences. CDH and I-CHASS have evaluated digital humanities workshops for over eight years collectively and will bring that experience to bear in constructing online evaluations to aid in the assessment of HpC. Data will be collected via surveys, content assessment instruments, and interviews over the course of the grant. Participants will be asked to assess the curriculum, a series evaluation to recommend strategies to improve the workshop sequence, and a final qualitative evaluation that will track the effects of this training on their individual digital humanities efforts. Program administrators at each center will also be surveyed to consider the ways in which the collaboration can be strengthened and changed over time to better serve the technological participants and their needs.

**Staff, Faculty, and Consultants (see Appendix III: Curriculum Vitae):  
University of South Carolina**

*Jennifer Guiliano* received a Bachelors of Arts in English and History from Miami University (2000), a Masters of Arts in History from Miami University (2002), and a Masters of Arts (2004) in American History from the University of Illinois before completing her Ph.D. in History at the University of Illinois (2010). She served as a Graduate Assistant and then a Post-Doctoral Research Assistant and Program Manager at the Institute for Computing in Humanities, Arts, and Social Sciences at the National Center for Supercomputing Applications from 2008- 2010 prior to assuming her role as Associate Director of the Center for Digital Humanities and a Research Assistant Professor at the University of South Carolina. Dr. Guiliano recently completed work as Primary Investigator on the NEH-funded: Sapheos Project, which explored digital collation and imaging. She currently serves as co-PI or contributor to a number of digital humanities projects including Cross-Disciplinary Investigations in Imaging and Image Analyses (an NSF-Sponsored Workshop in Information Integration and Informatics), the Handheld Technologies for Visual Culture Project with the Columbia Museum of Art, the Handheld Education for Public History project, and as a contributor to the Digging into Image Data to Answer Authorship Questions (sponsored by NEH, NSF, JISC, and SSRC) and the Rule of Law Collaborative (sponsored by the Department of Defense) projects. Dr. Guiliano will be responsible for supervision and management of the grant including implementing the projects governance, participant selection, curriculum and finance plans.

*Jijun Tang* received his M.S. (2002) and Ph.D. (2004) in Computer Science from the University of New Mexico after completing an M.S. (1996) in Offshore Engineering from Tianjin University (China) before joining the Department of Computer Science and Engineering at the University of South Carolina in 2004. He currently serves as a co-PI of an Office of Naval Research grant N00014-07-1-0686 to apply gaming paradigms and education into computer-aided design tools. He is the author of a number of articles on high performance computing and leads undergraduate and graduate training in humanities gaming. Dr. Tang will be responsible for the overall technical education and will lead the technical teams in assigning and coordinating resources for the humanities participants.

*Jun Zhou* completed her M.S. in Computer Science at Loyola University in 2002 before joining the Center for Digital Humanities as its Lead Programmer. Ms. Zhou is currently completing her Ph.D. in computer science while leading the digital humanities technical team. She serves as senior programmer on the Handheld Education for Public History project (an Android platform), the Handheld Technologies for Visual Culture Project (an iPhone platform), and the Paragon Project (a digital imaging project). Ms. Zhou will be responsible for providing technical consultations as well as working with the Primary Investigators to ensure adequate computational education for all participants.

**University of Illinois at Urbana-Champaign**

*Simon Appleford* received a Masters of Arts in Modern History and a Masters of Literature in Modern American History from the University of St. Andrews, Scotland, before joining NCSA in 2005. His interests in digital technologies and American history have led to several publications, including articles in CTWatch Quarterly and *Toward the Meeting of the Waters: Currents in the Civil Rights Movement in South Carolina* (University of South Carolina Press, 2007.) He was the principal organizer of e-Science for Arts and Humanities Research: Early Adopters Forum (2007), *Spatial Thinking in the Social Sciences and Humanities* (2006), and *Computing in Humanities, Arts, and Social Science* (2005). He is currently completing his PhD in History at the University of Illinois while serving as Project Manager at I-CHASS. He will be responsible for project coordination and collaboration between the computing centers and the humanities groups including the maintenance of the virtual community and the grant's findings.

*Peter Bajcsy*, Associate Director of I-CHASS, currently leads the Image Spatial Data Analysis group at the NCSA. He is also adjunct assistant professor in the Electrical and Computer Engineering Department and Computer Science Department at the University of Illinois. His research focuses on informatics and information technologies, and he has authored more than 16 papers in peer-reviewed journals that have been cited more than 200 times on scholar.google.com. He has co-authored 6 books or book chapters, and more than 75 conference papers, and has received more than \$3.3 million in research funding over the past 5 years as a principal or co-principal investigator. His work has been cited in MIT Technology Review, NCSA Headline News, HPCWire and NCSA Access. Dr. Bajcsy will be responsible for providing the large-scale imaging portions of the curriculum as well as deploying image tools for participants.

*Alan Craig*, Associate Director of Human-Computer Interaction at I-CHASS, has focused his career on the interface between humans and machines. He has been involved in many different capacities related to scientific visualization, virtual reality, data-mining, multi-modal representation of information, and collaborative systems during his career at the National Center for Supercomputing Applications where he has worked for the past 20 years. Dr. Craig is co-author of the book *Understanding Virtual Reality*, published by Morgan Kaufmann, and author of the book, *Developing Virtual Reality Applications*, also published by Morgan Kaufmann. Dr. Craig will be responsible for the augmented reality curriculum components and will work with participants in designing augmented reality tools for their usage.

*Marshall Scott Poole*, Director of I-CHASS, is a David and Margaret Romano Professorial Scholar, Professor in the Department of Communication, Senior Research Scientist at the National Center for Supercomputing Applications at the University of Illinois Urbana-Champaign, and Director of I-CHASS: The Institute for Computing in the Humanities, Arts, and Social Sciences. He received his Ph.D in 1980 from the University of Wisconsin-Madison. Scott has taught at the University of Illinois, the University of Minnesota, and Texas A&M University. His research interests include group and organizational communication, information systems, collaboration technologies, organizational innovation, and theory construction. He is the author of over 120 articles and book chapters. Current research interests include team behavior in massive multiplayer online games, structuring of communication and information technologies in groups, and the intersection of group and network theory in explaining large, dynamically changing groups and intergroup networks. Dr. Poole will be responsible for aiding participants in assessing their projects' collaboration networks and will aid them in utilizing the research networks established as part of this grant.

*Alex Yahja*, Assistant Director in Modeling at I-CHASS, works on the interface between technologies and the humanities, arts and social science as an Assistant Director for Modeling at I-CHASS. He received his PhD in computation, organizations and society and two Masters of Science degrees, one in engineering and public policy and one in robotics, from Carnegie Mellon University. His current research interests include spatial pattern learning, dynamic networks, and modeling and simulation. Dr. Yahja will be responsible for all social network curriculum and will also be responsible for designing/deploying social networking tools and analytics for participants.

**Budget Notes:**

HpC is requesting \$249,588 to fund two five-day-long residencies and a two-day-long culminating workshop that will serve thirty-five humanities participants nationally. This is in addition to ten computing center staff participants. Cumulatively, this grant will serve forty-five individuals, not including members of these scholars communities' who will benefit from the expertise brought back to their home institutions by participants. Of the thirty-five scholars, we will offer twenty-five participant slots for those who will participate in the full-sequence of training and five at-large positions at each workshop for local scholars interest in their local center's computational specialties.

Funding associated with participants costs were constructed using the following mathematical equations: for the five day workshop, Non-local participants will receive a \$500 travel allowance, a \$500 housing allowance, and \$45 per day in meals provided by the organizers. Local participants will receive a \$75 in meals provided by the organizers; for the two-day workshop, Non-local participants will receive a \$500 travel allowance, a \$200 housing allowance, and \$45 per day in meals provided by the organizers as well as an additional \$45 in reimbursable meals. Local participants will receive \$90 in meals provided by the organizers. The non-local participants are composed of the 25 selected humanities scholars and the five University of Illinois personnel. A complete budget and budget justification is available in Budget Justification and Budget.

**Citation:**

Peter Bajcsy, Jennifer Guiliano, and Kevin Hamilton. "Cross-Disciplinary Investigations in Image and Image Analysis: An NSF-Sponsored Workshop" National Science Foundation, Grant Submission, University of Illinois, Urbana, IL, 2009.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Illinois  
Institute for Computing in Humanities, Arts, and Social Science  
National Center for Supercomputing Applications

**Team members:**

University of Illinois  
Peter Bajcsy  
Jennifer Guiliano  
Kevin Hamilton

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Science Foundation.

**Citation:**

Peter Bajcsy, Jennifer Guiliano, and Kevin Hamilton. "Cross-Disciplinary Investigations in Image and Image Analysis: An NSF-Sponsored Workshop" National Science Foundation, Grant Submission, University of Illinois, Urbana, IL, 2009.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Illinois  
Institute for Computing in Humanities, Arts, and Social Science  
National Center for Supercomputing Applications

**Team members:**

University of Illinois  
Peter Bajcsy  
Jennifer Guiliano  
Kevin Hamilton

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Science Foundation.

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Peter Bajcsy

**Gender:**  Male  Female

**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)

American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)

Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Jennifer Guiliano

**Gender:**  Male  Female  
**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)  
 American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)  
 Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

**02 INFORMATION ABOUT PRINCIPAL INVESTIGATORS/PROJECT DIRECTORS(PI/PD) and  
co-PRINCIPAL INVESTIGATORS/co-PROJECT DIRECTORS**

---

Submit only ONE copy of this form for each PI/PD and co-PI/PD identified on the proposal. The form(s) should be attached to the original proposal as specified in GPG Section II.C.a. Submission of this information is voluntary and is not a precondition of award. This information will not be disclosed to external peer reviewers. **DO NOT INCLUDE THIS FORM WITH ANY OF THE OTHER COPIES OF YOUR PROPOSAL AS THIS MAY COMPROMISE THE CONFIDENTIALITY OF THE INFORMATION.**

---

**PI/PD Name:** Kevin Hamilton

**Gender:**  Male  Female  
**Ethnicity:** (Choose one response)  Hispanic or Latino  Not Hispanic or Latino

**Race:**  
(Select one or more)  
 American Indian or Alaska Native  
 Asian  
 Black or African American  
 Native Hawaiian or Other Pacific Islander  
 White

**Disability Status:**  
(Select one or more)  
 Hearing Impairment  
 Visual Impairment  
 Mobility/Orthopedic Impairment  
 Other  
 None

**Citizenship:** (Choose one)  U.S. Citizen  Permanent Resident  Other non-U.S. Citizen

**Check here if you do not wish to provide any or all of the above information (excluding PI/PD name):**

**REQUIRED: Check here if you are currently serving (or have previously served) as a PI, co-PI or PD on any federally funded project**

---

**Ethnicity Definition:**

**Hispanic or Latino.** A person of Mexican, Puerto Rican, Cuban, South or Central American, or other Spanish culture or origin, regardless of race.

**Race Definitions:**

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

**Black or African American.** A person having origins in any of the black racial groups of Africa.

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

---

**WHY THIS INFORMATION IS BEING REQUESTED:**

The Federal Government has a continuing commitment to monitor the operation of its review and award processes to identify and address any inequities based on gender, race, ethnicity, or disability of its proposed PIs/PDs. To gather information needed for this important task, the proposer should submit a single copy of this form for each identified PI/PD with each proposal. Submission of the requested information is voluntary and will not affect the organization's eligibility for an award. However, information not submitted will seriously undermine the statistical validity, and therefore the usefulness, of information received from others. Any individual not wishing to submit some or all the information should check the box provided for this purpose. (The exceptions are the PI/PD name and the information about prior Federal support, the last question above.)

Collection of this information is authorized by the NSF Act of 1950, as amended, 42 U.S.C. 1861, et seq. Demographic data allows NSF to gauge whether our programs and other opportunities in science and technology are fairly reaching and benefiting everyone regardless of demographic category; to ensure that those in under-represented groups have the same knowledge of and access to programs and other research and educational opportunities; and to assess involvement of international investigators in work supported by NSF. The information may be disclosed to government contractors, experts, volunteers and researchers to complete assigned work; and to other government agencies in order to coordinate and assess programs. The information may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records", 63 Federal Register 267 (January 5, 1998), and NSF-51, "Reviewer/Proposal File and Associated Records", 63 Federal Register 268 (January 5, 1998).

## List of Suggested Reviewers or Reviewers Not To Include (optional)

---

### **SUGGESTED REVIEWERS:**

Not Listed

### **REVIEWERS NOT TO INCLUDE:**

Not Listed

---



# CERTIFICATION PAGE

## Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the Authorized Organizational Representative or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, lobbying activities (see below), responsible conduct of research, nondiscrimination, and flood hazard insurance (when applicable) as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG) (NSF 10-1). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

## Conflict of Interest Certification

In addition, if the applicant institution employs more than fifty persons, by electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.A; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

## Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

## Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?

Yes

No

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

## Certification Regarding Lobbying

The following certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding \$100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding \$150,000.

## Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

- (1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.
- (2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.
- (3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

## Certification Regarding Nondiscrimination

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

## Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

- (1) community in which that area is located participates in the national flood insurance program; and
- (2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:

- (1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
- (2) for other NSF Grants when more than \$25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

## Certification Regarding Responsible Conduct of Research (RCR)

**(This certification is not applicable to proposals for conferences, symposia, and workshops.)**

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that, in accordance with the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.B., the institution has a plan in place to provide appropriate training and oversight in the responsible and ethical conduct of research to undergraduates, graduate students and postdoctoral researchers who will be supported by NSF to conduct research.

The undersigned shall require that the language of this certification be included in any award documents for all subawards at all tiers.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE		SIGNATURE		DATE	
NAME					
TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS			FAX NUMBER	

\* EAGER - EARly-concept Grants for Exploratory Research

\*\* RAPID - Grants for Rapid Response Research

# **Title: Cross-Disciplinary Investigations in Imaging and Image Analyses: An NSF-Sponsored Workshop**

Division: NSF CISE IIS –Information Integration and Informatics (III) program

Proposal Type: NSF 10-1 (General Solicitation)

Attention to: Steven Griffin

## **Project Summary**

From challenges related to data acquisition, storage, processing, and visualization to approaches to tool development, the potential for cross-fertilization among humanities, arts, social sciences and computational sciences are becoming more likely. No matter the domain or subfield, regardless of whether it is clinical and research domains utilizing medical imaging, universe telescope-based imaging, air-borne and ground imaging, satellite imaging, or historical documentary imaging, the challenges related to imaging and image analyses are manifest and common to all researchers and domain-specialists exploring imaging and imaging analysis.

Uniting the multi-disciplinary stakeholders in imaging and imaging analyses, this workshop presents a unique opportunity for cross-fertilizations, cross-collaboration, and information-exchange by using computational inquiries to bridge humanities, arts, and sciences research. Traditional workflows illustrate this compartmentalization and the need for cross-disciplinary exchanges. Scientists in humanities, social sciences and arts go from working in physical spaces to working in digital spaces, digitize historical artifacts via scanning, observe behavior via camera recordings and convey artistic perceptions via graphics. While this process is documented thoroughly, the resulting research generally only manifests in written publications or presentations targeted at their particular domain audience. Rarely do you find a humanities researcher reading the leading journals of computational science. Similarly, the physical sciences observe phenomena at micro and macro scales over space and time, undertake imaging and image analyses, and report their finding via prominent journals and computing organizations like the *International Journal of Computing* and the IEEE. They too are rarely engaged with humanities, and arts events and publications. This workshop presents the opportunity to elucidate connections between these communities via the imaging and image analyses. Significantly, the lack of forums for cross fertilization of imaging and image analysis efforts and ideas in multiple communities developing relevant technologies is not limited to the United States. Just as U.S. researchers and domain experts are seeking collaborative forums such as the one proposed here, so too are international researchers. As a result, this Workshop on Cross-Disciplinary Investigations into Imaging and Image Analyses will bring together representatives from US and foreign academic institutions interested in the use of imaging devices and applications of image analyses in their domains. In doing so, it will allow for global considerations of approaches to image and imaging analyses and will seed international collaborations for stakeholders across the humanities, arts, social sciences (HASS) and technology domains.

**Intellectual Merit:** The intellectual merit of Cross-Disciplinary Investigations in Imaging and Image Analyses lies in information exchange, cross-disciplinary collaboration, and cross fertilization of on-going and future efforts in the area of imaging and image analyses. For example, the computer science and electrical engineering communities have made significant advances in designing imaging techniques and image analysis algorithms while the HASS digital community has made strides in embracing new multi-media, social networking and web-based technologies. The workshop would create opportunities for the humanities, arts, and social science community to explore scientific instruments applied to imaging historical artifacts, as well as opportunities for the scientific community to investigate the adaptation of existing web 2.0 technologies to distributed research.

## TABLE OF CONTENTS

---

For font size and page formatting specifications, see GPG section II.B.2.

	<b>Total No. of Pages</b>	<b>Page No.* (Optional)*</b>
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	_____
Table of Contents	1	_____
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) <b>(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	8	_____
References Cited	1	_____
Biographical Sketches (Not to exceed 2 pages each)	6	_____
Budget (Plus up to 3 pages of budget justification)	3	_____
Current and Pending Support	5	_____
Facilities, Equipment and Other Resources	3	_____
Special Information/Other Supplementary Docs/Mentoring Plan	0	_____
Appendix (List below. ) <b>(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)</b>	_____	_____
Appendix Items:		

\*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

---

## Project Description

The University of Illinois, including the National Center for Supercomputing Applications (NCSA), the School of Art and Design, and the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), propose to organize a workshop on Cross-Disciplinary Investigation in Imaging and Image Analyses that will facilitate education, training and informational exchange among multiple scientific disciplines including humanities, arts, social science, science, and computational researchers. Organized collaboratively between NCSA, the School of Art and Design, and I-CHASS, this workshop will be uniquely positioned to draw from the varied scientific disciplines listed above. NCSA and I-CHASS have well-documented relationships with their disciplinary user and research-communities and have staged highly successful workshops to seed cross-fertilization, collaboration, and community exchange. Recent workshops and conferences that would be built upon for attendee and presenters include, but are not limited to: Imaging and Image Analyses Applied to Historical Objects (April 2009: 86 participants); Humanities, Arts, and Sciences Technology Advanced Collaboratory annual meetings (April 2009: 205 participants), High Performance Computing: Bio-Informatics Workshop for SC08 (July 2008: 45 participants); Imaging at Illinois at Beckman (October 2009: 29 presenters). Each of these workshops privileged interdisciplinary approaches to humanities, arts, and social science research that demonstrated the importance of advanced computing approaches. The School of Art and Design has attracted many students that cut across humanities, arts and social sciences, and would be well positioned to lead the educational aspects of the proposed workshop.

With such diverse disciplinary communities, this workshop will bring together representatives from US and foreign academic institutions engaged in imaging and imaging analyses. The workshop envisions large-group keynote presentations as well as break-out sessions. Possible topics for inclusion are listed below. We anticipate an audience that will consist of domestic and international researchers as well as University of Illinois students. The students will learn about tools useful for studying imaging and image analyses while the researchers will define joint projects. Before the workshop we will circulate proposal solicitations to encourage participants to team up at the meeting.

### *The need for this workshop*

From challenges related to data acquisition, storage, processing, and visualization to approaches to tool development, the potential for cross-fertilization among humanities, arts, social sciences and the sciences is just beginning to be explored. No matter the domain or subfield, regardless of whether it is clinical and research domains utilizing medical imaging, universe telescope-based imaging, air-borne and ground imaging, satellite imaging, or historical documentary imaging, the challenges related to imaging and image analyses are manifest and common to all researchers and domain-specialists exploring imaging and imaging analysis.

Uniting the multi-disciplinary stakeholders in imaging and imaging analyses, this workshop presents a unique opportunity for cross-fertilizations, cross-collaboration, and information-enchange by using computational inquiries to bridge humanities, arts, and sciences research. Traditional workflows illustrate this compartmentalization and the need for cross-disciplinary exchanges. Scientists in humanities, social sciences and arts go from working in physical spaces to working in digital spaces, digitize historical artifacts via scanning, observe behavior via camera recordings and convey artistic perceptions via graphics. While this process is documented thoroughly, the resulting research generally only manifests in written publications or presentations targeted at their particular domain audience. Rarely do you find a humanities researcher reading the leading journals of computational science. Similarly, the physical sciences

observe phenomena at micro and macro scales over space and time, undertake imaging and image analyses, and report their finding via prominent journals and computing organizations like the *International Journal of Computing* and the IEEE. They too are rarely engaged with humanities, and arts events and publications. This workshop presents the opportunity to elucidate connections between these communities via the imaging and image analyses. Significantly, the lack of forums for cross fertilization of imaging and image analysis efforts and ideas in multiple communities developing relevant technologies is not limited to the United States. Just as U.S. researchers and domain experts are seeking collaborative forums such as the one proposed here, so too are international researchers. As a result, this Workshop on Cross-Disciplinary Investigations into Imaging and Image Analyses will bring together representatives from US and foreign academic institutions interested in the use of 2D and 3D imaging devices and applications of image and video analyses in their domains. In doing so, it will allow for global considerations of approaches to image and imaging analyses and will seed international collaborations for stakeholders across the humanities, arts, social science, science, and technology domains.

### **Workshop theme**

Our intention is to illustrate approaches to imaging and image analyses from the diverse communities of humanities, arts, social science, and computational researchers who would be attending this workshop. In doing so, the pedagogical goal of the workshop will be to understand the challenges associated with imaging and image analyses that common across disciplinary fields, to develop collaborations among user and research-communities, and to cross-fertilize these communities with ideas and strategies from related imaging and image analyses projects. The workshop will emphasize strategies for collaboration and will devote significant attention to development between these communities. While it is not trivial to build such cross-disciplinary bridges, we are confident, given the wide-ranging interest of these communities in developing new approaches and tools for imaging and image analyses, that the workshop these of cross-disciplinary investigations and collaborations via imaging and image analyses will be highly successful.

### **Preliminary list of workshop topics**

We have carefully selected eight topics equally split between HASS researchers focusing on traditional historical materials and new media materials. The topics related to traditional historical materials include 2D scanning and 3D imaging, as well as analyses of 2D scans and 3D historical artifacts. The topics related to new media materials (primarily born digitally) include content-based image or video retrieval, preservation and esthetics of imagery from cognitive, technological and cultural perspectives. In order to cover many perspectives on these topics, we tried to select two speakers per topic that would represent the problem domain and technology perspectives.

Topic:	Speakers and Topics
Imaging: 2D scanning technology	(1) Bob Markley (English Department, UIUC) – Topic: Digitization and optical character recognition of the 18th-century connect manuscripts (2) Richard Marciano (RENCI, University of North Carolina); Topic: T-RACES: Testbed for the Redlining Archives of California's Exclusionary Spaces  Substitute: Valerie Hotchins (UIUC Library); Topic: Digitization

	and study of palimpsests in Rare book collections
Image Analyses: 2D scans	<p>(1) Mike Meredith, University of Sheffield, UK; Topic: Web-based dissemination of large image collections via Virtual Vellum (<a href="http://www.shef.ac.uk/french/research/froissart/vvellum.htm">http://www.shef.ac.uk/french/research/froissart/vvellum.htm</a>)</p> <p>(2) Mara R. Wade (Germanic Languages and Literatures, UIUC) – Renaissance Emblem Books</p> <p>Substitute: Anne D. Hedeman (Art History and Medieval Studies UIUC); Topic: Analyses of Illustrations in Froissart and Shrewsbury manuscripts</p>
Imaging: 3D imaging	<p>(1) Wayne Pittard ( Spurlock Museum, UIUC); Topic: InscriptiFact: A Networked Database of Ancient Near Eastern Inscriptions Project</p> <p>(2) Todd Presner (University of California at Los Angeles); Topic: The use of historical images for building HyperCities: Berlin and Los Angeles</p> <p>Substitute: Bruce Zuckerman (University of Southern California): Topic: InscriptiFact:</p>
Image Analyses: 3D content	<p>(1) John Bonnett (Brock University, CA), Topic: 3D Virtual Buildings Project, <a href="http://www.brocku.ca/history/faculty/jbonnett/index.php">http://www.brocku.ca/history/faculty/jbonnett/index.php</a></p> <p>(2) Lisa M. Snyder (the UCLA urban simulation team), Topic: The interactive computer model of the World’s Columbian Exposition of 1893</p> <p>Substitute: Geraldine Heng (or Ana Boa-Ventura) (University of Texas), Topic: Clash of Civilizations</p>
Image Analyses: Content-Based Image Retrieval	<p>(1) Dean Rehberger (or Justine Richardson)) (MSU); Topic: Understanding 19-20th century quilt imagery; Quilt Photographs (<a href="http://www.quiltindex.org/">http://www.quiltindex.org/</a>)</p> <p>(2) James Wang/ Jia Li (Penn State) Topic: Intelligent media annotations</p> <p>Substitute: Eamonn Keogh, University of California- Riverside, Topic: AI for Understanding Cultural Heritage:</p>
Video analyses: Content-Based Video Retrieval	<p>(1) Sang-Wook Lee; (South Korea, Sogang University), Topic: video in-painting</p> <p>(2) David Forsyth, (CS UIUC); Topic: Building models of animals from video</p> <p>Substitute: Scott Poole, (Communication Dept. UIUC): Topic: Comparison of skype and 3D video based communication</p>
Image and Video Analyses: Preservation	<p>(1) Karen Cariani (WGBH); Topic: media archive migrations</p> <p>(2) Marc Downie, (MIT’s Media Lab): Topic: 3D geometry of a location from photographs taken from diverse angles</p> <p>Substitute: Peter Bajcsy, (NCSA/UIUC): Topic: preservation and</p>

	access by keeping file formats current
Esthetics of imagery from cognitive, technological and cultural perspectives	<p>(1) Jer Thorpe or M.Simon Levin, (the Surrey Art Gallery's TechLab, CA): Topic: Glocal (global + local) collaborative and multifaceted digital art project</p> <p>(2) Lisa Nakamora (UIUC); Topic: social behavior and gaming (<a href="http://en.wikipedia.org/wiki/Lisa_Nakamura">http://en.wikipedia.org/wiki/Lisa_Nakamura</a>)</p> <p>Substitute: Kevin Hamilton (UIUC, School of Art and Design), Topic: interpretation of mass production and consumption of imagery</p>

### *Additional potential workshop speakers*

While we will approach the speakers selected above, we have also created a list of additional potential workshop speakers. The final list of workshop speakers would be determined based on the availability of identified candidates as well as from a widely-distributed solicitation of attendees.

#### ***Speakers presenting domain problems:***

- (1) Linne Mooney (University of York, UK ) - Computer Identification of Medieval Scribal Handwriting from the Medieval English Palaeography perspective
- (2) Will Noel (Walters Art Museum in Baltimore); Topic: Multi-spectral Analyses of Archimedes palimpsests
- (3) Pat Seed (University of California at Irvine ); Topic: The Development of Mapping on the West and South Coasts of Africa by Portuguese Navigators and Cartographers from 1434-1504
- (4) Paul Lovejoy (York University, Canada ); Topic: Maps for understanding the African diaspora
- (5) Valerie Hotchins (UIUC Library); Topic: Digitization and study of palimpsests in Rare book collections
- (6) Bonnie Mak (GSLIS UIUC); Topic: Image preservation in the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project
- (7) Max Edelson (History Department, University of Virginia) - The History of Colonial British America and the Atlantic World from Historical Maps
- (8) Susan Noakes (University of Minnesota); Topic: Digitization of Italian & French Medieval Manuscripts
- (9) Ray Kea, (History, UC Riverside); Topic: Analyses of pre-colonial West African and Ghanaian economic, cultural, and social history from historical maps.
- (10) Geraldine Heng (English, University of Texas, Austin); Topic: Studies of English medieval manuscripts
- (11) Virginia Khun (University of Southern California, Institute for Multimedia Literacy); Topic: Multimedia analyses
- (12) Nick Barbules (Education policy studies, UIUC); Topic: Ubiquitous learning using multimedia
- (13) Geraldine Heng (or Ana Boa-Ventura) (University of Texas), Topic: Clash of Civilizations
- (14) Jonathan Fineberg (UIUC), Topic: Film making and preservation
- (15) Bill Seaman, Duke University, Topic: media-oriented poetics

- (16) Nancy Turner, associate conservator in the department of paper conservation at the J. Paul Getty Museum ;Topic: pigment analysis and its applications to understandings of workshop practice

***Speakers presenting information technology requirements and prototype solutions:***

- (3) Roger Easton, (Rochester Institute of Technology); Topic: spectral analyses of Archimedes palimpsests
- (4) John Daugman (Cambridge, UK); Topic: Computer Identification of Medieval Scribal Hands
- (5) Ana Boa-Ventura (Communications, University of Texas, Austin); Topic: Images in Virtual worlds
- (6) James Gain (Department of Computer Science, University of Cape Town); Topic: Visualization for the Humanities
- (7) Jose Castro (Costa Rica, CeNAT); Topic: Cluster computing for the Humanities
- (8) Bruno Schultze (Brazil, LNCC); Topic: Grid computing for the Humanities
- (9) Sang-Chul Lee (South Korea, Inha University); Topic: document analyses
- (10) Hany Farid (Dartmouth University), Topic: forensic image analyses
- (11) Michael Toth (Toth associates): Topic: Hyperspectral imaging of Palimpsest
- (12) William H. Mischo (UIUC library), Topic: Preservation of images and multimedia files
- (13) David Stork (Ricoh corporation), Topic: Geometry based modeling of art

***Recent meetings on the same subject, including dates and locations***

There have been efforts over the past two to three years to bring together researchers from the Humanities, Arts, and Social Sciences domains together with the computer scientists and information technologists. These efforts led to projects such as HASTAC [1] and Bamboo [2], as well as to several workshops in which we have participated [3-4]. Additionally, this effort builds on the highly successful NSF sponsored workshop Imaging and Image Analyses Applied to Historical Analyses [5]. Imaging and Image Analyses Applied to Historical Analyses was co-directed by Peter Bajcsy of I-CHASS/NCSA, Anne D. Hedeman, and Karen Fresco of the University of Illinois. Designed to facilitate education, training and information exchange among multiple scientific disciplines, the Imaging and Image Analyses workshop brought together representatives from academic institutions in the United States and abroad and from US museums. Humanists, social scientists, and artists were paired with computer scientists at the workshop in order to present complementary views on topics related to imaging and image analyses of historical objects. The intent of the workshop was to examine the process of going from actual physical objects to digital objects made available via the Internet and the related process of enabling computer assisted learning over large digital collections for education and research. The overarching goal of the workshop was to understand the challenges associated with imaging and image analyses that are inherent in this process, as well as solutions, needs and opportunities for further research. Specific topics covered included: stroke analysis of paintings, historical mapping, emblem books, quilt imagery, multimedia analyses, palimpsest, Virtual Vellum, and manuscripts from the 18th and 19th centuries. Attended by 86 participants, this highly successful workshop generated the formulation of four new research groups on Image and Imaging Analyses as well as the establishment of an Imaging Analyses network for humanities, arts, social science, and technology researchers. This network and these research groups will be leveraged as part of the Cross-Disciplinary Investigations in Imaging and Image Analyses Workshop.

Other conferences have recognized the computational needs for humanities and included sessions on digital humanities, for instance, the eScience conference [6] attended by us most recently. Other international eScience conferences such as the meetings in UK [7] and in Korea [8] have also included digital humanities. The proposed workshop would extend these previously organized efforts but would be much more focused on specific imaging technologies and the related image analysis problems. The uniqueness of the proposed workshop lies in its more focused technical scope while simultaneously bridging the communities of humanities, arts, social science, biomedical, geo-spatial, and computational science.

### ***Workshop organization***

The workshop will run during two days at National Center for Supercomputing Applications and at the University of Illinois. NCSA classrooms can accommodate 60 people for small breakout sessions. Additionally, there are 30-60 person classrooms available at the Siebel Computer Science Building next door. All rooms are equipped with audio-visual equipment and with power extensions for laptops. Auditoriums at NCSA can accommodate 200 people and will allow for simultaneous broadcasting of keynote topics to the web and/or additional locations if needed.

In order to attract a large number of community representatives and aim at the broadest impact, the proposed workshop would be organized by the National Center for Supercomputing Applications and the Institute for Computing in Humanities, Arts, and Social Science at the University. We anticipate each organization leading the involvement of its particular domain-community (e.g. NCSA leading computational sciences, and I-CHASS leading humanities, arts, and social science disciplinary recruitment.) We anticipate 150-200 attendees from the humanities, arts, science and computational community. The funds requested as part of this proposal will be used to support the organization and logistics associated with the humanities, arts, social science and computational community as well as participant costs for those communities.

### ***The names of the chairperson and members of organizing committees and their organizational affiliations:***

Peter Bajcsy, NCSA & I-CHASS, UIUC  
Kevin Hamilton, School of Art and Design, UIUC  
Jennifer Guiliano , I-CHASS, UIUC

### ***Information on the location and probable date(s) of the meeting***

NCSA/UIUC; depending on the funding availability either October – November 2010 or March – April 2011.

### ***Soliciting Workshop Participants and Presenters***

#### ***The method of announcement or invitation:***

We would use the following methods

- Email invitations to individual speakers
- Classroom announcements for University of Illinois students as well as e-mail lists for appropriate University of Illinois groups.

- Web page posting at the NCSA and I-CHASS web pages
- Distribution via mailing lists, professional organization announcement lists and social media outlets

The workshop is expected to have more than 50% participants and speakers from groups that are underrepresented in science and engineering (e.g., underrepresented minorities, women, and persons with disabilities).

To solicit these participants and presenters, the Cross-Disciplinary Investigations in Imaging and Image Analyses Workshop organizers will utilize a custom-designed online submission tool. Built for the Institute for Computing in Humanities, Arts, and Social Science in partnership with the National University Community Research Institute, the online submission tool allows for the web-based collection of data for conference participants including, but not limited to: registration of participants, gathering of biographical statements, resumes, and contact information, submission of presentation abstracts, papers, and other pertinent publications. Additionally, the online submission tool includes a parallel review process that allows conference organizers to: 1) establish panels of reviewers of all conference submissions, 2) conduct a review process using pre-established criteria for selection, 3) a mathematical ordering of conducted reviews that ranks submissions and generates an ordered list of presenters based on their review scores, 4) contact all reviewers and applicants with results, and 5) import/export materials including submissions and reviewer comments in .doc, .pdf, .txt, and .html formats. All of the participant and reviewer functionalities are accessible via an easy-to-use web-based interface. This system has been successfully used by I-CHASS for the 1,000,000 CPU Competition (Spring 2009), the CONICIT-Costa Rica Competition (Fall 2009), HASTAC 2010 (Spring 2010) and will be used in the forthcoming 2,000,000 CPU Competition.

### *Education Activities*

We would like to pursue two educational aims of the proposed workshop. First, it is the workshop attendance and exposure of students to a variety of imaging and image analysis problems. Second, it is the engagement of the students in the follow-up classes and independent studies that would leverage their knowledge gained during the workshop and would entice the students to pursue the hands-on experience with imaging and image analyses.

We would accomplish the first aim by advertizing at UIUC, and the second aim by offering independent studies for students interested in some topics (advertized during the workshop) and by drafting courses like "Computer-aided Photographic Research: Developing Critical Frameworks Through Image Analysis" to enrich the students interested in learning more about the workshop topics. The UIUC units where we plan to advertize include Architecture, Art and Design, Landscape Architecture, Graduate School of Library Information Sciences (GSLIS), Computer Science, Electrical Engineering, Advertising, Journalism, Urban Planning and Media Studies.

### *Dissemination of workshop results*

Findings from this project will be disseminated in traditional and innovative ways. We will encourage and maintain collaboration among participants by creating a web-portal that allows electronic dissemination and maintains a constant web-based presence. The primary goal of the web portal will be to disseminate the experience, presentations and learned lessons. The

secondary goal will be facilitated by the inclusion of web modules embedded in the portal that allows activities tracking (provenance data), collaboration support (including blogs, chat, and wikis), and social networking support. Creating a multi-way networked activity centered on the digital humanities and lessons learned from each project will be disseminated to the wider humanities, academic audience and the general public through our virtual community in addition to the more traditional online papers, journal articles and research reports. Our ambition is to achieve continual and energetic discussion and collaboration for each group and as a collaborative.

The two-day workshop, which will be webcast, will be summarized in a final report. The report will outline long-range technological planning goals that can be undertaken by the partnership of each humanities group with computer scientists. In serving not just invited participants but also scholars and students interested in imaging and imaging analyses, the workshop should achieve the objective to engage in collaboration with colleagues from diverse backgrounds.

We would also like to conduct an evaluation of the workshop via evaluation tools provided by I-CHASS. I-CHASS has evaluated workshops and conferences of a cross-disciplinary nature before and our internal resources would collect data via surveys, content assessment instruments, and interviews over the course of the workshop.

### *Intellectual Merit*

The intellectual merit of Cross-Disciplinary Investigations in Imaging and Imaging Analyses lies in information exchange, cross-disciplinary collaboration, and cross fertilization of on-going and future efforts in the area of imaging and image analyses. For example, the computer science and electrical engineering communities have made significant advances in designing imaging techniques and image analysis algorithms while the HASS digital community has made strides in embracing new multi-media, social networking and web-based technologies. The workshop would create opportunities for the humanities, arts, and social science community to explore scientific instruments applied to imaging historical artifacts, as well as opportunities for the scientific community to investigate the adaptation of existing web 2.0 technologies to distributed research.

### *Broader Impact*

The broader impacts resulting from the proposed workshop are in crossing boundaries of humanities, arts, social science, and computational science disciplines in the applications of imaging and multi-media technologies, cross pollination of ideas on the use of custom developed imaging instruments to other domain applications, and identifying joint research efforts and educational engagement of students in using imaging and image analysis tools.

## References:

- [1] HASTAC (Humanities, Arts, Science, and Technology Advanced Collaboratory) project; sponsored by the McArthur Foundation; URL: <http://www.hastac.org/about>
- [2] Bamboo project; Project Bamboo is an 18-month planning and community design program; URL: <http://projectbamboo.uchicago.edu/about-us>.
- [3] Summer workshop on Computational Methods in the Humanities, Arts, and Social Sciences, August 2007, SDSC, San Diego, CA, <http://www.sc-education.org/wiki/index.php/UCIHRI07#Schedule>
- [4] Summer workshop on High Performance Computing in the Humanities, Arts, and Social Sciences: Information-Rich Environments for Research and Teaching, July 28, 2008 –August 2, 2008, NCSA, Urbana, IL, [http://www.chass.uiuc.edu/Events/Entries/2008/7/27\\_High\\_Performance\\_Computing\\_in\\_the\\_Humanities%2C\\_Arts%2C\\_and\\_Social\\_Sciences%3A\\_Information-Rich\\_Environments\\_for\\_Research\\_and\\_Teaching.html](http://www.chass.uiuc.edu/Events/Entries/2008/7/27_High_Performance_Computing_in_the_Humanities%2C_Arts%2C_and_Social_Sciences%3A_Information-Rich_Environments_for_Research_and_Teaching.html)
- [5] Imaging and Image Analyses Applied to Historical Objects, April 22-23, 2009, University of Illinois. NSF Sponsored.
- [6] The 4th IEEE International Conference on e-Science, December 8-12, 2008, Indianapolis, IN
- [7] The UK eScience All-Hands Meeting, Crossing Boundaries: Computational Science, E-Science and Global E-Infrastructures, 8th - 11th September 2008, Edinburgh, United Kingdom, <http://www.nesc.ac.uk/talks/>.
- [8] The Korea eScience All Hands Meeting, September 8-9, Daejeon Convention Center, Daejeon, Korea, <http://www.ahm.or.kr/>.

## Peter Bajcsy

### a. Professional Preparation.

Slovak Technical University, Bratislava, Czechoslovakia

Major: Technical Cybernetics and Measurement Techniques

Degree & Year: Diploma Engineer, 1987

University of Pennsylvania, Philadelphia, PA

Major: Electrical Engineering

Degree & Year: Master of Science, 1994

University of Illinois, Urbana-Champaign, IL

Major: Electrical and Computer Engineering

Degree & Year: Doctor of Philosophy, 1997

### b. Appointments.

2007 – Present	Associate Director, ICHASS, Illinois Informatics Institute, UIUC
2003 – Present	Adjunct Assistant Professor ECE, UIUC, IL
2002 – Present	Adjunct Assistant Professor CS, UIUC, IL
2001 – Present	Research Scientist NCSA, UIUC, IL
1998 – 2001	Senior Scientist SAIC/DEMACO, Inc., Champaign, IL.
1997 - 1998	Senior Software Engineer Cognex_Corporation, Acumen Products Group, Portland, OR

### c. Publications. (5 Most Closely Related to Proposed Research)

- (1) Bajcsy P., Y-F. Lin, A. Yahja, and C-Y. Kim, "A Framework for Accurate Geospatial Modeling using Image Ranking and Machine Learning," Journal of Hydroinformatics (accepted January 2010).
- (2) McFadden W., R. Kooper, S-C. Lee, and P. Bajcsy, "Comprehensive and Scalable Appraisals of Contemporary Documents," a book chapter in the book on Machine Learning", by IN-TECH – the free online access publishing, ISBN 978-953-7619-X-X, April 2010
- (3) Singh A., B. S. Minsker, and P. Bajcsy, "Image-Based Machine Learning For Reduction Of User-Fatigue In An Interactive Model Calibration System," Journal of Computing in Civil Engineering, 1, 6 (2009), DOI:10.1061/(ASCE)CP.1943-5487.0000026
- (4) Kooper R., A. Shirk, S-C. Lee, A. Lin, R. Folberg and P. Bajcsy, "3D Volume Reconstruction Using Web Services," Elsevier International Journal on Computers in Biology and Medicine, Vol. 38/4, pp 490-500; DOI information: 10.1016/j.compbimed.2008.01.015, 2008
- (5) Lee S-C. and P. Bajcsy, "Three-dimensional Volume Reconstruction Based on Trajectory Fusion from Confocal Laser Scanning Microscope Images," Journal of Computer Vision and Image Understanding 110 (2008) 19–31.

### (Other Significant Publications and Software Systems.)

- (1) Peter Bajcsy and Maryam Moslemi, "Discovering Salient Characteristics of Authors of Art Works," IS&T/SPIE Electronic Imaging, 17 - 21 January 2010, San Jose Convention Center, Section - Computer Vision and Image Analysis of Art, Paper 7531-10 presented on January 18th at 1:20pm
- (2) Malik R. and P. Bajcsy, "Achieving Color Constancy Across Multiple Cameras", ACM International Conference on Multimedia, Beijing, China, October 19 - 24, 2009 (~ 30% acceptance).
- (3) Bajcsy P., K. McHenry, H-J Na, R. Malik, A. Spencer, S-K. Lee, R. Kooper, and M. Frogley, "Immersive Environments For Rehabilitation Activities," ACM International Conference on Multimedia, Beijing, China, October 19 - 24, 2009 (~ 27.5% acceptance).
- (4) Malik R. and P. Bajcsy, "Automated Placement of Multiple Stereo Cameras," OMNIVIS'2008, the Eighth Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras held in conjunction with ECCV 2008, Marseille, France, October 12-18th, 2008

- (5) Bajcsy P., S-C. Lee, A. Lin and R. Folberg, "3D Volume Reconstruction of Extracellular Matrix Proteins in Uveal Melanoma from Fluorescent Confocal Laser Scanning Microscope Images," Journal of Microscopy, Vol. 221, Pt 1 January 2006, pp. 30–45

**d. Synergistic Activities.**

- Reviewer for journals and conferences: Asian Conference on Computer Vision (ACCV 09), International Conference on Machine Learning and Applications (ICMLA 07,08,09), Pattern Recognition (07, 08), IEEE Transactions on Information Forensics and Security (07, 08, 09), Bioinformatics (07, 09), IEEE Signal Processing Letters (07), EURASIP Journal on Applied Signal Processing (06), Journal of Microscopy (06, 08), IEEE Trans. on Intelligent Transportation (07), IEEE Transactions on Medical Imaging (06, 09), IEEE Transactions on Geoscience and Remote Sensing (05, 06, 07), IEEE Geoscience and Remote Sensing Letters (08), and IEEE on Pattern Analysis and Machine Intelligence (04).
- Served on the NIST advisory board for the SHIELD project 03-04; on the NSF Information Technology Research (ITR) Review Panels for Division of Informative Biology and Division of Information & Intelligent Systems, 04 (twice), 05 (once), 09 (once), 10 (twice).
- Served as an associate editor the ACM Transactions on Multimedia Computing Communications and Applications 2010, and a program committee member on the IEEE Workshop on Computer Vision Methods for Bioinformatics (in conjunctions with IEEE CVPR 05), and on the International Conference on Machine Learning and Applications (ICMLA 07 & 08).
- Contributions to the science of learning; Collaborates with humanists, artists, social scientists, earth scientists, engineers, citizens with disabilities, clinicians in hospitals, biologists, neuroscientists, psychologists and veterinary medicine experts on X-informatics problems.

**e. Collaborators & Other Affiliations**

**(i) Collaborators.**

Scott Poole, Department of Communication, UIUC, IL

Anne D. Hedeman, School of Arts and Design, UIUC, IL

Klara Nahrstedt, Department of Computer Science, UIUC, IL

Amelia Bartholomew, Dept of Surgery, University of Illinois, Chicago, IL;

Barbara Minsker, Department of Civil and Environmental Engineering, UIUC, IL.

Yu-Feng Lin, Illinois State Water Survey, Champaign, IL;

Jiawei Han, Department of Computer Science, UIUC, IL;

Robert Folberg, Oakland University William Beaumont School of Medicine, Rochester, MI

**(ii) Graduate and Postdoctoral Advisors.**

MS Thesis Advisor: Sallem Kassam, Department of Electrical Engineering, UPENN, Philadelphia, PA

PhD Thesis Advisor: Narendra Ahuja, Department of Electrical and Computer Engineering, UIUC, IL

**(iii) Thesis Advisor and Postgraduate-Scholar Sponsor.**

**Completed/Current Ph.D. – 1/2; M.S. – 8/0, undergraduate students 20/1**

Peter Groves (M.S. 12/2003), Technical Consultant at Moiret, Chicago, IL; Sunayana Saha (M.S.

12/2003), Citadel Investment Group, Chicago, IL.; David Scherba (M.S. 05/2005), Qualcomm, Denver,

CO.; Wei-Wen Feng (M.S. 05/2006), PhD, UIUC; Sang-Chul Lee (PhD 05/2006), Professor, Inha

University, Korea; Miles Johnson (M.S. 12/2007), PhD student, UIUC; Shadi Ashnai (M.S. 12/2007),

Wolfram Research, Champaign, IL; Chandra Ramachandran (M.S. 12/2009), Qualcomm, CA.

**Jennifer E. Guiliano, Ph.D.**

Post-Doctoral Research Assistant  
Institute for Computing in Humanities, Arts, and Social Science  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801  
217-369-8355  
guiliano@illinois.edu

**Professional Preparation**

Miami University, History and English Literature, *B.A.*, 2000  
Miami University, History, *M.A.*, 2002  
University of Illinois, History, *M.A.*, 2004  
University of Illinois, History, *Ph.D.*, 2010

**Academic Experience**

Post Doctoral Research Assistant, Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), National Center for Supercomputing Applications, January 16, 2010- July 15, 2010.  
Graduate Assistant, Institute for Computing in the Humanities, Arts, and Social Science (I-CHASS), National Center for Supercomputing Applications, January 2008- January 2010.  
Lecturer, Department of History, History 273: Illinois History, June 2008 to August 2008.  
Graduate Teaching Assistant, Department of History, August 2007 to December 2007, August 2003 to August 2004, August 2002 to December 2002.  
Graduate Research Assistant, Department of History, August 2006 to December 2006, May 2003- May 2008.  
Lecturer, Department of History, History 172/173: History of the United States, 1877 to the Present, June 2006 to August 2006.  
Graduate Assistant, Department of History, January 2005 to May 2006, August 2004 to May 2005  
Graduate Assistant, American Indians of Illinois Online, Dr. Brenda Farnell, Department of Anthropology, June 2003 to August 2003.  
Graduate Assistant, Native American House, December 2002 to May 2003.

**Honors, Awards, Notable Achievements**

*Graduate Student Essay Award*, North American Society for Sport History, Thirtieth Annual Conference, French Lick, IN, May 2002.  
University of Illinois at Urbana-Champaign student recipient of the 2003 Committee on Institutional Cooperation selective seminar on "Nineteenth-Century Native American Representations," at the D'Arcy McNickle Center for American Indian Studies at the Newberry Library, under the direction of Susan Sleeper-Smith, Michigan State University.  
*Incomplete List of Teaching Excellence*, Fall 2002, Fall 2007, Summer 2008. University of Illinois at Urbana-Champaign.  
Department of History Fellowship, University of Illinois at Urbana-Champaign, August 2005 to December 2005.  
Department of History Fellowship, University of Illinois at Urbana-Champaign, January 2007 to May 2007.

**Selected Publications**

"Creating Tradition and Defining "Indianness", *Studies in Symbolic Interaction* 34 (2010)(London: Emerald Group Publishing, 2010).  
"Chasing Objectivity? Critical Reflections on History, Identity, and the Public Performance of Indian

Mascots.” (Publication Pending in *International Review of Qualitative Research*, special issue on Sport and Auto-Ethnography, edited by Michael Giardina.)

Co-author, Test Bank and Instructor’s Resource Manual, in *The People: A History of Native America*, authored by Dr. Frederick Hoxie, R. David Edmunds, and Neil Salisbury.

“Questioning Indians on the Field: Native Americans, Sports Mascots, and the Education of Whiteness.” (Publication Pending in *The Journal of Sport and Social Issues*, edited by C.L. Cole and Adrian Burgos.)

### **Collaborators and Affiliations**

Simon Appleford, Institute for Computing in Humanities, Arts, and Social Science

Peter Bajcy, National Center for Supercomputing Applications

Alan Craig, Institute for Computing in Humanities, Arts, and Social Science

Kevin Franklin, Institute for Computing in Humanities, Arts, and Social Science

Kevin Hamilton, University of Illinois

David Miller, University of South Carolina

Marshall Scott Poole, Institute for Computing in Humanities, Arts, and Social Science

Alex Yahja, Institute for Computing in Humanities, Arts, and Social Science

## Kevin Hamilton

### a. Professional Preparation.

Rhode Island School of Design, Providence RI

Major: Painting

Degree & Year: Bachelor of Fine Arts, 1996

Massachusetts Institute of Technology

Major: Visual Studies

Degree & Year: Master of Science, 2000

### b. Appointments.

2006 – Present Program Chair, New Media, School of Art and Design, UIUC, IL  
2002 – Present Associate Professor, New Media, School of Art and Design, UIUC, IL  
2001 – 2002 Visiting Instructor, Art Program, Grand Valley State University, MI  
2000 – 2001 Adjunct Instructor, Gordon College, MA  
1999 – 2002 Summer Instructor, Rhode Island School of Design, Providence, RI

### c. Publications. (5 Most Closely Related to Proposed Research)

- (1) Hamilton K., N. O’Gorman, "Nuclear Memory at the Interface," *Sighting Memory: The Intersection of Visual Practices and Practices of Memory*, Anne Demo and Brad Vivian ed. (Forthcoming book chapter.)
- (2) Hamilton K., L. Nakamura, "Return of the Digital Native: Interfaces, access and racial difference in *District 9*," *FlowTV* v. 11.12, March 2010.
- (3) Hamilton K., N. O’Gorman, "At the Interface: The Loaded Rhetorical Gestures of Nuclear Legitimacy and Illegitimacy," co-authored with Ned O’Gorman. Forthcoming in *Communication and Critical/Cultural Studies*.
- (4) Sharmin, Bailey, Coats, Hamilton, "Understanding Knowledge Management Practices for Early Design Activity," *Proceedings of the 2009 ACM/Human-Computer Interaction Conference*
- (5) Adamczyk P, K.Hamilton "Urban Computing and Mobile Devices," *IEEE Distributed Systems Online*, vol. 8, no. 7, 2007.

### (Recent Art Exhibitions)

- (1) "BCL/IGB." Illinois State Arts commission, Institute for Genomic Biology, UIUC (budget: \$35,000).
- (2) "I will lift up mine eyes unto the hills." Performance for DIRT Festival, Links Hall Chicago.
- (3) "On Location." *Blackbird* online journal of literature and the arts, Virginia Commonwealth University.
- (4) "Chronozone 4." project for Mobile Studios in Bratislava, Budapest and Sofia.
- (5) "Department of Rhythmanalysis." Gahlberg Gallery, College of DuPage. Glen Ellyn, IL.

### (Recent Awards and Fellowships)

- (1) NEH/Vectors Fellowship, IML University of Southern California (Summer 2010)
- (2) Fellow, Cornell University Society for the Humanities (2009-10, declined).
- (3) *Empyre* workshop scholarship, Anderson Ranch Arts Center, Colorado, 2008.
- (4) *Reference Check* research residency, Banff New Media Institute, Calgary, Canada, 2007.
- (5) Illinois Arts Council Fellowship - Interdisciplinary Category, 2007. (\$7000)

### d. Synergistic Activities.

- Workshop Co-organizer, *HCI and New Media: Evaluation + Methodology*. Association of Computing Machinery / Computer-Human Interaction Conference, San Jose, CA. 2007.
- Workshop Co-organizer, *Tools in Support of Creative Collaboration*. Association of Computing Machinery / Creativity and Cognition Conference, Washington, DC. 2007.

- Symposium Co-Coordinator, *Walking as Knowing as Making: A Peripatetic Examination of Place*. University of Illinois. 2006
- Contributions to New Media Education: Conducts workshops and lectures on visual imaging for Research Groups in the sciences, humanities and social sciences. Developed and administers a new undergraduate program in New Media Art.
- Contributions to interdisciplinary scholarship: Has organized graduate-level courses on collaborative cross-disciplinary design problems. Studied creativity and conducted surveys of practitioners in the arts and sciences.

#### **e. Collaborators & Other Affiliations**

##### **(i) Collaborators.**

Weimo Zhu, Department of Kinesiology, UIUC, IL  
 Ned O’Gorman, Department of Communication, UIUC, IL  
 Brian Bailey, Department of Computer Science, UIUC, IL  
 Piotr Adamczyk, Metropolitan Museum of Art, New York  
 M.Simon Levin, University of British Columbia, Vancouver.

##### **(ii) Graduate and Postdoctoral Advisors.**

n/a

##### **(iii) Thesis Advisor and Postgraduate-Scholar Sponsor.**

**Completed/Current Ph.D. – 1/0; M.F.A. – 15/0**

Celiany Rivera, PhD Candidate, Institute of Communication Research  
 MFA New Media:

Bobby Belote *2010*  
 Ashwyn Collins *2010*  
 Heather Ault *2010*  
 Skot Wiedmann *2009*  
 Collin Bradford *2008*  
 Matthew Yapchaian *2007*  
 Rose Marshack *2005*  
 Katerie Gladdys *2005*

MFA Sculpture:

Katie Latona *2010*  
 Karin Hodgins Jones *2008*  
 Jennifer Danos *2004*

MFA Painting:

J. Meredith Warner *2005*  
 David Prinsen *2004*  
 Sarah Kanouse *2004*  
 Anna Callahan *2003*

# SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION <b>University of Illinois at Urbana-Champaign</b>				FOR NSF USE ONLY			
				PROPOSAL NO.	DURATION (months)		
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Peter Bajcsy</b>				AWARD NO.	Proposed	Granted	
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				NSF Funded Person-months		Funds Requested By proposer	Funds granted by NSF (if different)
				CAL	ACAD	SUMR	
1. <b>Peter Bajcsy - PI</b>				0.20	0.00	0.00	\$ 1,540
2. <b>Jennifer Guiliano - Co-PI</b>				0.20	0.00	0.00	846
3. <b>Kevin Hamilton - Co-PI</b>				0.00	0.00	0.20	1,336
4.							
5.							
6. ( 0 ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00	0
7. ( 3 ) TOTAL SENIOR PERSONNEL (1 - 6)				0.40	0.00	0.20	3,722
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)							
1. ( 0 ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00	0
2. ( 1 ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.20	0.00	0.00	854
3. ( 0 ) GRADUATE STUDENTS							0
4. ( 0 ) UNDERGRADUATE STUDENTS							0
5. ( 0 ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)							0
6. ( 0 ) OTHER							0
TOTAL SALARIES AND WAGES (A + B)							4,576
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)							1,505
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)							6,081
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)							
TOTAL EQUIPMENT							0
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)							0
2. FOREIGN							0
F. PARTICIPANT SUPPORT COSTS							
1. STIPENDS \$ 3,000							
2. TRAVEL 15,566							
3. SUBSISTENCE 6,000							
4. OTHER 6,000							
TOTAL NUMBER OF PARTICIPANTS ( 150 )							
TOTAL PARTICIPANT COSTS							30,566
G. OTHER DIRECT COSTS							
1. MATERIALS AND SUPPLIES							2,250
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION							1,250
3. CONSULTANT SERVICES							2,500
4. COMPUTER SERVICES							180
5. SUBAWARDS							0
6. OTHER							0
TOTAL OTHER DIRECT COSTS							6,180
H. TOTAL DIRECT COSTS (A THROUGH G)							42,827
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)							
<b>MTDC (Rate: 58.5000, Base: 12261)</b>							
TOTAL INDIRECT COSTS (F&A)							7,173
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)							50,000
K. RESIDUAL FUNDS							0
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)							\$ 50,000
M. COST SHARING PROPOSED LEVEL \$ 0				AGREED LEVEL IF DIFFERENT \$			
PI/PI NAME <b>Peter Bajcsy</b>				FOR NSF USE ONLY			
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION			
				Date Checked	Date Of Rate Sheet	Initials - ORG	

# SUMMARY PROPOSAL BUDGET Cumulative

ORGANIZATION <b>University of Illinois at Urbana-Champaign</b>				FOR NSF USE ONLY		
				PROPOSAL NO.	DURATION (months)	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR <b>Peter Bajcsy</b>				AWARD NO.	Proposed	Granted
					NSF Funded Person-months	
A. SENIOR PERSONNEL: PI/PI, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)				CAL	ACAD	SUMR
1. <b>Peter Bajcsy - PI</b>				0.20	0.00	0.00
2. <b>Jennifer Guiliano - Co-PI</b>				0.20	0.00	0.00
3. <b>Kevin Hamilton - Co-PI</b>				0.00	0.00	0.20
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)				0.00	0.00	0.00
7. ( <b>3</b> ) TOTAL SENIOR PERSONNEL (1 - 6)				0.40	0.00	0.20
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. ( <b>0</b> ) POST DOCTORAL SCHOLARS				0.00	0.00	0.00
2. ( <b>1</b> ) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)				0.20	0.00	0.00
3. ( <b>0</b> ) GRADUATE STUDENTS						
4. ( <b>0</b> ) UNDERGRADUATE STUDENTS						
5. ( <b>0</b> ) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)						
6. ( <b>0</b> ) OTHER						
TOTAL SALARIES AND WAGES (A + B)						<b>1,540</b>
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)						<b>846</b>
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)						<b>1,336</b>
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING \$5,000.)						
TOTAL EQUIPMENT						<b>0</b>
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)						<b>0</b>
2. FOREIGN						<b>0</b>
F. PARTICIPANT SUPPORT COSTS						
1. STIPENDS \$ <b>3,000</b>						
2. TRAVEL <b>15,566</b>						
3. SUBSISTENCE <b>6,000</b>						
4. OTHER <b>6,000</b>						
TOTAL NUMBER OF PARTICIPANTS ( <b>150</b> )						
TOTAL PARTICIPANT COSTS						<b>30,566</b>
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES						<b>2,250</b>
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION						<b>1,250</b>
3. CONSULTANT SERVICES						<b>2,500</b>
4. COMPUTER SERVICES						<b>180</b>
5. SUBAWARDS						<b>0</b>
6. OTHER						<b>0</b>
TOTAL OTHER DIRECT COSTS						<b>6,180</b>
H. TOTAL DIRECT COSTS (A THROUGH G)						<b>42,827</b>
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)						<b>7,173</b>
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)						<b>50,000</b>
K. RESIDUAL FUNDS						<b>0</b>
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)						<b>\$ 50,000</b>
M. COST SHARING PROPOSED LEVEL \$ <b>0</b>				AGREED LEVEL IF DIFFERENT \$		
PI/PI NAME <b>Peter Bajcsy</b>				FOR NSF USE ONLY		
ORG. REP. NAME*				INDIRECT COST RATE VERIFICATION		
				Date Checked	Date Of Rate Sheet	Initials - ORG

C \*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

BUDGET JUSTIFICATION

Principal Investigator: Peter Bajcsy

Period: 9/1/10 – 8/31/10

- |  | Project Dollars |
|--|-----------------|
| <b>A. Senior Personnel</b>   | <b>\$3,722</b>  |
| Salary support of 0.20 months is requested for the PI, Peter Bajcsy, & Co-PI's, Jennifer Guiliano & Kevin Hamilton to prepare, execute and document the workshop. Salaries are based on actual UIUC AY2010 rates and are incremented at a rate of 3.0% each year.  |                 |
| <b>B. Other Personnel</b>  | <b>\$854</b>    |
| Salary support of 0.20 months is requested for project manager Simon Appleford to assist with multimedia needs, ordering food, and travel and hotel arrangements. Salaries are based on actual UIUC AY2010 rates and are incremented at a rate of 3.0% each year.  |                 |
| <b>C. Fringe Benefits</b>  | <b>\$1,505</b>  |
| Fringe benefits are charged at a rate of 32.88% on faculty and postdoc salaries. Benefits include retirement, worker's compensation, health, life and dental insurance, termination, and Medicare. Fringe benefits are charged at a rate of 4.49% on graduate student salaries. Benefits include worker's compensation and health, life and dental insurance.  |                 |
| <b>F. Participant Costs/Workshop</b>   | <b>\$30,566</b> |
| For the two day workshop, we anticipate between 60 to 120 participants based on last year's attendance figures. The selected invited speakers listed in the proposal would be reimbursed for their travel costs (up to \$750 for domestic and \$1250 for international travelers) and we would provide 3 nights hotel for the invited speakers. During the workshop, all registered participants would receive workshop materials, refreshments during the breaks and boxed lunches. |                 |
| <b>G. Other Direct Costs</b>   | <b>\$6,180</b>  |
| <u>Materials and Supplies:</u> Funds budgeted for materials and supplies will cover the expendable supplies and equipment needed to conduct the research program and include items that are normally required to operate a research program.   |                 |
| <u>Publication Costs:</u> This includes the estimated cost of preparing and publishing project results.  |                 |
| <u>Consultant Services:</u> Funds are requested to support the participation of Registration System to manage abstract submissions, review process and electronic posting of workshop information.   |                 |
| <u>Computer Services:</u> NCSA assesses a standard fee to support infrastructure support costs and retrieval of scientific and technical information. The established computer service rate at NCSA is \$225.00 per month per employee. The NCSA infrastructure support fee is currently under review by campus officials.   |                 |
| <b>I. Indirect Costs</b>   | <b>\$7,173</b>  |
| Indirect costs are assessed at a rate of 58.5% of Modified Total Direct Costs (MTDC). MTDC is direct costs less equipment, tuition remission, and subawards in excess of \$25,000.   |                 |

## Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.

Investigator: Peter Bajcsy (Page 1 of 3)	Other agencies (including NSF) to which this proposal has been/will be submitted. None
--	---

Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: CDI-Type II: Sensor Web Workbench for Coupled Human-Natural Systems  Source of Support: NSF Total Award Amount: _____ Total Award Period Covered: 9/1/10-8/31/14 Location of Project: University of Illinois Urbana Champaign Person-Months Per Year Committed to the Project. _____ Cal: 3.0 Acad: _____ Sumr: _____
--

Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Collaborative Research: CI-TEAM Implementation Project Contributing to Peace and Non-Violence through Classroom and Service Learning-based Cyberinfrastructure Source of Support: NSF Collaborative Total Award Amount: \$344,149 Total Award Period Covered: 1/1/2011 – 12/31/2012 Location of Project: University of Illinois at Urbana-Champaign Person-Months Per Year Committed to the Project. _____ Cal: 0.5 Acad: _____ Sumr: _____
--

Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: HCC: Collaborative Research: PHYSNET: Physical Interaction Using the Internet Source of Support: National Science Foundation Total Award Amount: \$900,000 Total Award Period Covered: 2/1/07 – 1/31/10 Location of Project: University of Illinois Person-Months Per Year Committed to the Project. _____ Cal: 0.25 Acad: _____ Sumr: _____
---

Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Cross-Disciplinary Investigations in Imaging and Image Analyses: An NSF-Sponsored Workshop Source of Support: NSF EAGER Total Award Amount: \$50,000 Total Award Period Covered: 09/01/2010 - 08/31/2011 Location of Project: University of Illinois at Urbana-Champaign Person-Months Per Year Committed to the Project. _____ Cal: 0.2 Acad: _____ Sumr: _____
---

Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Digging into image data to answer authorship related questions  Source of Support: NSF Total Award Amount: \$100,000 Total Award Period Covered: 01/01/2010 till 03/30/2011 Location of Project: UIUC Person-Months Per Year Committed to the Project. _____ Cal: 0.1 Acad: _____ Sumr: _____
---

Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support Project/Proposal Title: Collaborative Research: CDI-Type II: Groupscope: Instrumenting Research on Interaction Networks in Complex Social Contexts; Source of Support: NSF Total Award Amount: \$1,761,568 Total Award Period Covered: 1/1/2010-12/31/2014 Location of Project: NCSA/UIUC Person-Months Per Year Committed to the Project. _____ Cal: 2.0 Acad: _____ Sumr: _____
--

\*If this project has previously been funded by





## Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.				
Investigator: Guiliano, Jennifer			Other agencies (including NSF) to which this proposal has been/will be submitted.	
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support				
Project/Proposal Title:				
<b>Digging into Data to Answer Authorship Related Questions</b>				
Source of Support: NSF				
Total Award Amount: \$100,000		Total Award Period Covered: 01/01/2010 till 03/30/2011		
Location of Project: University of Illinois at Urbana-Champaign				
Person-Months Per Year Committed to the Project.    7.6		Cal: 1.0	Acad:	Sumr:
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Resubmission Planned in Near Future <input type="checkbox"/> *Transfer of Support				
Project/Proposal Title:				
Cross-Disciplinary Investigations in Imaging and Image Analyses: An NSF-Sponsored Workshop				
Source of Support: NSF EAGER				
Total Award Amount: \$50,000		Total Award Period Covered: 9/1/2010 – 8/31/2011		
Location of Project: University of Illinois at Urbana-Champaign				
Person-Months Per Year Committed to the Project.		Cal: 0.2	Acad:	Sumr:
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support				
Project/Proposal Title:				
Source of Support:				
Total Award Amount:		Total Award Period Covered:		
Location of Project:				
Person-Months Per Year Committed to the Project.		Cal:	Acad:	Sumr:
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Resubmission Planned in Near Future <input type="checkbox"/> *Transfer of Support				
Source of Support:				
Total Award Amount:		Total Award Period Covered:		
Location of Project:				
Person-Months Per Year Committed to the Project.		Cal:	Acad:	Sumr:
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support				
Source of Support:				
Total Award Amount:		Total Award Period Covered:		
Location of Project:				
Person-Months Per Year Committed to the Project.		Cal:	Acad:	Sumr:
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.				



## Current and Pending Support

(See GPG Section II.D.8 for guidance on information to include on this form.)

The following information should be provided for each investigator and other senior personnel. Failure to provide this information may delay consideration of this proposal.			
Investigator: Hamilton, Kevin	Other agencies (including NSF) to which this proposal has been/will be submitted.		
Support: <input type="checkbox"/> Current <input checked="" type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: <b>Cross-Disciplinary Investigations in Imaging and Image Analyses</b>			
Source of Support: NSF			
Total Award Amount: \$50,000		Total Award Period Covered: 9/1/2010-8/31/2011	
Location of Project: University of Illinois at Urbana-Champaign			
Person-Months Per Year Committed to the Project.		Cal:	Acad:
		Sumr:	0.20
Support: <input checked="" type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Resubmission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title: <b>SoD-TEAM: Developing Computational Tools that Facilitate Individual and Group Creativity</b>			
Source of Support: NSF			
Total Award Amount: \$742,405		Total Award Period Covered: July 27,2006 – July 31, 2010	
Location of Project: University of Illinois, Urbana-Champaign			
Person-Months Per Year Committed to the Project.		Cal:	Acad:
		Sumr:	
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project: University of Illinois at Urbana-Champaign			
Person-Months Per Year Committed to the Project.		Cal:	Acad:
		Sumr:	
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Resubmission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project:			
Person-Months Per Year Committed to the Project.		Cal:	Acad:
		Sumr:	
Support: <input type="checkbox"/> Current <input type="checkbox"/> Pending <input type="checkbox"/> Submission Planned in Near Future <input type="checkbox"/> *Transfer of Support			
Project/Proposal Title:			
Source of Support:			
Total Award Amount:		Total Award Period Covered:	
Location of Project:			
Person-Months Per Year Committed to the Project.		Cal:	Acad:
		Sumr:	
*If this project has previously been funded by another agency, please list and furnish information for immediately preceding funding period.			



## **L. Facilities, Equipment and Other Resources**

NCSA continues to support user communities by offering the resources that are the foundations of advanced cyberinfrastructure. The total computational resources exceed 43 TF supported by over 1 PB of disk storage as part of the infrastructure. The systems are on an internal 10GbE network. Below is a summary of those resources.

### **NCSA Compute Resources**

#### *Dell Intel® 64 Tesla Cluster [lincoln]*

Lincoln consists of 192 compute nodes (Dell PowerEdge 1950 dual-socket nodes with quad-core Intel Harpertown 2.33GHz processors and 16GB of memory) and 96 NVIDIA Tesla S1070 accelerator units. Each Tesla unit provides 345.6 gigaflops of double-precision performance and 16GB of memory.

#### *Dell Intel® 64 Cluster [abe]*

Peak performance: 89.47 TF (62.68 TF sustained)

Top 500 list debut: #8 (June 2007)

This Dell blade system has 1,200 PowerEdge 1955 dual socket, quad core compute blades, an InfiniBand interconnect and 400 TB of storage in a Lustre filesystem. Abe is a shared resource that is 60% allocated through the National Science Foundation allocation process, with the remaining time allocated at the discretion of the NCSA leadership to serve state of Illinois, University of Illinois strategic initiatives, and NCSA's Private Sector Program Partners.

#### *SGI Altix [cobalt]*

1,024 Intel Itanium 2 processors

Peak performance: 6.55 TF (6.1 TF sustained)

Top 500 list debut: #48 (June 2005)

Updated peak performance August 2008: 8.2 TF

The SGI Altix consists of several Intel Itanium 2 processor shared-memory systems running the Linux operating system. Note: This resource will be retired on March 31 2010.

#### *IBM IA-64 Linux Cluster [mercury]*

1,774 Intel Itanium 2 1.3/1.5 GHz processors, 4 GB and 12 GB memory/node

Peak performance: 10.23 TF (7.22 TF sustained)

Top 500 list debut: #15 (June 2004)

The IA-64 TeraGrid cluster consists of 887 IBM dual processor Itanium 2 nodes, running SuSE Linux and Myricom's Myrinet cluster interconnect network. Note: This resource will be retired on March 31 2010.

#### *Mass Storage System (MSS)*

NCSA's hierarchical archival storage system is available for permanent storage of data. Access is via the FTP and SSH based transfer clients, including GridFTP clients. NCSA's mass storage now holds more than six petabytes of data and has the capacity to archive ten petabytes of data.

## **Infrastructure SAN**

284 TB of SAN connected storage for infrastructure and special projects.

## **High-Performance Network**

All computing platforms are interconnected to a multi-10gigabit network core. The NCSA high-performance computing environment has access to the Abilene high-performance network through a shared 10-gigabit-per-second connection. NCSA also is one of the leading sites for I-WIRE, an optical networking project funded by the state of Illinois. I-WIRE provides lambda services for several projects, including NCSA's 30-gigabit-per-second connection to the TeraGrid network.

## **Display Systems**

*Tiled Display Wall:* This environment consists of 40 NEC VT540 projectors, arranged in a matrix 5 high and 8 across. The output of the NEC VT540s is rear-projected towards a single screen, creating a large-format, high-resolution image space that is 8192 x 3840 pixels. A 40-node PC Linux cluster is used to drive the display wall. The machines are dual-processor Intel Xeons, running at 2.4 GHz, with Nvidia FX 5800 Ultra graphics accelerator cards, and communicating over Myrinet.

*High Definition Passive Stereo Theater:* The NCSA High Definition Passive Stereo Theater is a 1920x1080 display on an 6' x 3'5" screen. The projectors used are JVCD-1LA. The display is driven by a dual AMD Opteron 242 processor running at 1.6 GHz. Graphics hardware consists of a Nvidia Quadro FX3000.

## **Applications Software**

NCSA offers a variety of third-party applications and community codes that are installed on the high-performance systems at NCSA. These applications cover a wide range of science and engineering domains, data analytics and visualization, mathematics and statistics. Complete information on the packages available and detailed descriptions of them are available at:

<http://hpcsoftware.ncsa.uiuc.edu/Software/user/index.php?view=NCSA>.

## **Additional sensor and imaging laboratory equipment in the image spatial data analysis (ISDA) group**

ISDA NCSA also operated a variety of sensor and imaging devices for laboratory experiments. The list of the equipment is provided below:

- Hyperspectral camera by Opto-Knowledge Systems Inc.,
- Multispectral camera MS2150 by DuncanTech,
- Thermal infrared cameras, model Omega manufactured by Indigo Systems Corporation,
- Network Color Camera SNC-RZ30N PTZ Pan/Tilt/Zoom by Sony,
- Wireless miniature color cameras by Samsung,
- Wireless MICA sensors with multiple configurations of sensor boards by Crossbow Inc.,
- Pioneer 2 robot (P2 DX8) by ActivMedia Robotics,
- AmigoBot by Media Robotics
- Wireless audio sensors by Audio-technica Corp.,
- IS-300 Pro Precision Motion Tracker by InterSense,

- Bright field microscope by Bausch & Lomb with Sony camera,
- JTAG in-circuit emulator for testing MICA sensors.
- Multiple clusters of Point Grey visible spectrum cameras that are connected to quad core PCs to achieve real-time 3D reconstruction of a scene.
- 52" LCD displays
- Several tripods and TV cart to enable portable deployment of cameras
- Two wii controls for view control in 3D virtual space.
- TYZX stereo cameras

The School of Art and Design at University of Illinois offers the campus a base of operations for instruction in digital media. Courses and workshops in screen-based media take place in state-of-the-art instructional labs that are open and available to students 14 hours a day. Students can choose from seven different Apple Computer labs running a full suite of Adobe and Apple products for multimedia authoring, including one lab running Windows environments for 3D graphics and one lab set up for video editing, with dual monitors and DV decks for acquisition. Other digitization tools offered include multiple standard digital scanners, three film scanners, one large flatbed scanner, and four Cintiq stylus-screens. Students can output their work using a host of Epson inkjet printers, printing in widths up to 48 inches and lengths of their choosing. The 3D output lab also features a Universal Laser Systems X-660 laser-cutter and FDM Stratasys Dimension SST 3-D printer. Students taking classes in these facilities also have access to a circulating pool of equipment for recording and capture of visual and audio data, including a range of digital SLR cameras for still imagery and digital video cameras for high and low-end applications. Tripods, microphones and lighting equipment are also available. Courses taught in the labs are augmented by lectures in one of four "smart rooms" within the School of Art and Design, which allow for networked presentations and projections in multiple media formats.

## National Digital Stewardship Residency Workplan

### Stage One (month 1-2):

<b>Resident Tasks:</b>	<b>Interdependent Mentor Tasks:</b>
<p>State of the Art Research Report for access models and collection interfaces for born-digital literary materials</p> <ul style="list-style-type: none"> <li>Undertake qualitative survey of reference services by libraries providing access to born-digital collections</li> <li>Undertake qualitative survey collection interfaces provided by libraries holding born-digital collections</li> <li>Compile survey results into draft state of the art research report</li> </ul>	<ul style="list-style-type: none"> <li>Use professional networks to connect fellow with leading research libraries working with born-digital collections for access models and collection interfaces survey</li> <li>Provide feedback on draft state of the art research report</li> </ul>
<p>Deliverable: Completed state of the art research survey on access models and collection interfaces for born-digital materials</p>	

### Stage 2 (months 2-4)

<p>Education/training in user-centered design</p> <ul style="list-style-type: none"> <li>complete reading list as generated by mentors and HCIL on user-centered design</li> <li>complete reading list on requirements analysis and rapid prototyping</li> </ul>	<p>Provide access to:</p> <ul style="list-style-type: none"> <li>Reading list in user-centered design</li> <li>HCIL faculty for training in user-centered design</li> <li>MITH and ITD staff for training in rapid prototyping and development</li> </ul>
<p>Deliverable: Presentation to BDWG on user-centered design principles and methodologies; plan for organizing requirement-gathering workshops</p>	

### Stage 3 (months 4-6)

<p>Requirements gathering workshop</p> <ul style="list-style-type: none"> <li>conduct workshop with Special Collections and archives staff</li> <li>conduct workshop with electronic literature researchers and digital humanists</li> </ul> <p>Shadow special collections staff assisting patrons in reading room; collect field observations of current practice</p> <p>Draft recommendations for access policies, reference services, and on-site researcher support for born-digital collections</p>	<p>Arrange meetings:</p> <ul style="list-style-type: none"> <li>with Special Collections and archives staff for requirements workshop</li> <li>with ITD regarding current library technologies</li> <li>with electronic literature researchers and digital humanists regarding access needs</li> <li>with born-digital working group regarding access needs</li> </ul> <p>Provide comments on recommendations draft</p>
<p>Deliverable: Completed recommendations for access policies, reference, and on-site researcher support for born-digital collections</p>	

**Stage 4 (months 6-8)**

Iterative prototyping of online collection interfaces with supervision from Libraries ITD and MITH Complete lo-fi paper mock-ups of potential interfaces Design wireframes and interaction behaviors (page flows, transitions, navigation) Collaborate with BDWG and ITD to implement working prototype (time permitting)	Ensure Fellow's access to digital objects and appropriate technical resources including software licenses and hardware Provide comments on lo-fi paper mock-ups, wireframes and interaction behaviors Collaborate on working prototype
Deliverable: Finalized design brief including wireframes and interaction design	

**Stage 5 (months 8-9):**

Draft final report and present results to BDWG	Collaborate with the Fellow to prepare and submit a presentation to an appropriate professional conference or publication on their Residency research Present results from Residency to leadership in the University Libraries and the College of Arts and Humanities
Final report to include: State of the art research survey on access models for born-digital collections Access model for UMD for born-digital materials Access interface prototype for UMD for born-digital materials	

The Fellow will gain demonstrable experience with reference models, user-centered design, and prototyping. As more and more institutions add born-digital materials to their collections, they will need individuals capable of developing and integrating policies and access models where none existed previously. This residency provides the necessary background to successfully articulate the issues surrounding the access of born-digital archival collections and the expertise to provide solutions to this problem.

# The Palmetto History Project

## Table of Contents

<b>Narrative</b> .....	1
Nature of the Request.....	1
Humanities Content.....	2
Project Format and Participant Experience.....	11
Project Resources.....	12
Project History .....	14
Audience, Marketing, and Promotion .....	14
Evaluation, Final Product, and Dissemination.....	15
Organizational Profiles .....	15
Project Team.....	16
Work Plan .....	17
Fundraising Plan.....	18
<b>Work Samples for Digital Media Components</b> .....	19
<b>Bibliography</b> .....	20
<b>Resumes and Letters of Commitment</b> .....	22
<b>Budget</b>	

# The Palmetto History Project

## Nature of the Request

Clemson University has undertaken a collaboration between public historians, computer scientists, and heritage tourism sites to develop *The Palmetto History Project*—an initiative that seeks to leverage new developments in mobile app development to bring the rich history of Upcountry South Carolina to a wide audience that includes both site visitors and “virtual tourists.” Although the Upcountry has a rich and diverse history, the region is often overshadowed by other parts of South Carolina and remains underrepresented in historical scholarship. A primary goal of *The Palmetto History Project*, therefore, is to correct this historiographical gap and reveal to scholars, lay historians, and the general public the critical role played by the region’s social, economic, and political history in shaping and reflecting larger southern and national trends in American History. There is a growing interest amongst scholars and those involved in heritage tourism to make southern history and historical sites more accessible and engaging to both students and the general public. Indeed, the increasingly ubiquitous nature of cell phones capable of delivering information- and media-rich experiences has created an unparalleled opportunity for public historians and heritage tourism sites to engage with visitors and the public in innovative ways. To leverage this confluence, work has already begun on the development of a user-friendly mobile application that will allow users to explore the rich tapestry of South Carolina’s heritage, placing its history in broader southern, national, and global contexts. *The Palmetto History Project* will bring together historians, students, technology experts, and heritage site directors that empowers users to explore historical sites in innovative ways whether they are at home sitting in front of a computer, traveling through the state, or physically at the site. These digital tools help orient different kinds of learners—kinetic, visual, and aural—in a way that will reach out to a diverse audience. *Palmetto History* will help public historians engage with academics to provide a model for historic sites interested in using mobile applications as a means of attracting and engaging tourists. We will help sites become “participatory museums” that serve their communities, educate visitors, and facilitate ongoing conversations about Upcountry South Carolina’s history.

The Palmetto History Project has been fortunate to receive limited start-up funds from Clemson University—a \$5,000 grant from the Clemson CyberInstitute and \$1,500 from the Clemson University Service Alliance—that has allowed the PIs to collaborate on the creation of a prototype mobile application (designed for use with Apple’s iPad), with content from five nearby heritage tourism sites. To ensure that the project continues to meet the needs of users, public historians, and heritage tourism sites, the project requires a more substantial investment of start-up funds, which will be used to continue the development of the prototype application, incorporate materials from additional heritage tourism sites located within the Upcountry, and to convene an advisory board of experts in the history of the region to advise on the focus of the historical content to be incorporated into the project.

## Humanities Content

The Upcountry of South Carolina, defined by historian Lacy Ford as the region bounded by the “Sand Hills” break to the East and the borders of both North Carolina and Georgia to the North and West,<sup>1</sup> is especially rich in opportunities to explore southern history. Indeed, the Upcountry is home to a startling number of heritage tourism sites that preserve and interpret a broad range of historical moments and themes in South Carolina’s history. Despite these strong historical connections across the region, scholars tend to neglect the Upcountry when studying the social, political, and economic history of South Carolina, which leads to a skewed understanding of the importance of the state’s role in both southern and national history.

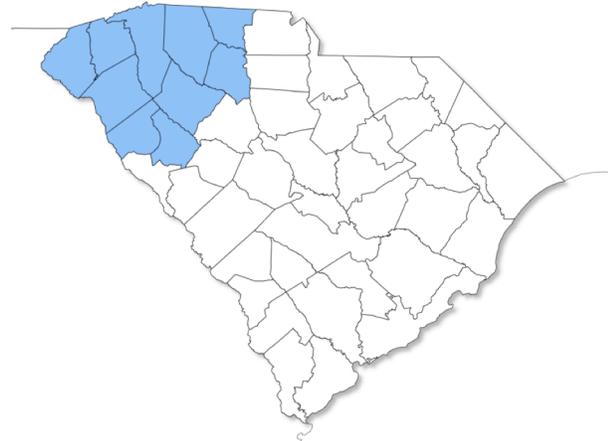


Figure 1: The Upcountry of South Carolina

Even so, a review of the scant literature of the region shows the Upcountry emerging as an arbiter of southern trends. From its earliest colonial period, when settlers entered a world dominated by the powerful Cherokee nation and fought as “Regulators” to establish law and order on the frontier, to the lynching of Willie Earle in 1947, which garnered national attention and prompted the state to pass an anti-lynching law, the Upcountry defined itself as a region of conflict, change, and violence.

Historians have long noted the frontier as a place of contestation, and this is particularly true of the Upcountry. Lightly settled by whites who generally moved down the Great Wagon Road throughout the colonial period, the region’s Upper Piedmont was the home of Keowee Town and other prominent settlements of the Cherokee Nation. The region was engulfed in the Cherokee War from 1760-1 that resulted in the Cherokees’ loss of land and the movement of settlers into the region. Historians note the difference between the white settlers of the Upcountry and those of the more prosperous lower regions; primarily Scots-Irish and German, these settlers had fewer slaves, generally less arable land, and a religious outlook that differed greatly from the settlers of the Low Country. This was the region in which evangelical change swept the districts, and this was the region in which the Regulators fought for a judicial infrastructure that the Charleston political elite had not supplied until the late 1760s. As historian Walter Edgar notes, it was this population that was skeptical of the patriot elites in Charleston at first, until the British began to threaten their land, their families, and their way of life. It was then that these men formed militia units and began to rout the British army, first at Huck’s Defeat, then at King’s Mountain, and finally in Cowpens and Ninety-Six. Initially reluctant to enter the fray, upcountry men became the critical linchpin in turning the tide of the American Revolution, a fact often overlooked by historians

---

<sup>1</sup> Lacy K. Ford, *Origins of Southern Radicalism*, viii.

who pay more attention to the battles fought in New Jersey, Massachusetts, New York, and Virginia.<sup>2</sup>

Sites that detail this challenging period of Upcountry history are surprisingly numerous. We seek to eventually include the following sites on our app to highlight the importance of the region's history:

- Currently a state park, *Oconee Station* is the only surviving fort and trading post that served as a critical trading center for the European Americans in the 1790s. It has two original structures, both of which are on the national register of historic places.
- *King's Mountain National Military Park*, *Cowpens National Battlefield*, and *Ninety-Six National Historic Site* provide interpretations of revolutionary history, including living history encampments, guided tours, and cinematic and/or museum exhibits.
- *Walnut Grove Plantation*, built in the 1750s, interprets colonial life, slavery, and revolutionary history on-site.
- *Historic Brattonsville* includes structures built from the mid-1700s through the mid-nineteenth century and, though living history and tours, interprets the events leading to Huck's Defeat, the lives of the wealthy planters in the Upper Piedmont through the Bratton family, and the lives of the slaves the Brattons owned.

Perhaps historians of the early Republic and antebellum periods overlook the Upcountry because the opulent plantations of the Low Country provide such a stark contrast to northern industry. In fact, one can argue that the Low Country of South Carolina is more an example of either southern exceptionalism at the most, or at least as more representative of the large plantations of the new frontiers of Mississippi and Alabama. In fact, according to Lacy K. Ford, the upcountry more resembled southern states such as North Carolina, Virginia, and Tennessee, with its split between the Appalachian Upper Piedmont and the lower Piedmont, in which the majority of the black population of the Upcountry was found. Because of its vast diversity of geography, the Upcountry featured everything from large plantations that produced cotton as cash crops to small yeoman farms that kept to the motto of "subsistence first" and produced just a few bales of cotton each year, at least until the 1840s.<sup>3</sup> In addition, the Upcountry featured one of the earliest "retreats" for

---

<sup>2</sup> Tom Hatley describes white/Cherokee interactions from the mid-1600s through the end of the Revolution, arguing that the encounters between Cherokee and white men and women were critically important the formation of each identity's culture on the frontier in *The Dividing Paths*. Walter Edgar's *Partisans and Redcoats* also covers this period in history, from the settlement of the area by immigrants traveling down the Great Wagon Road through the Appalachians, to the Cherokee War that resulted in a land grab in the Upcountry, to the move from support of the British to support of the Revolution after Cherokees attacked frontier settlements. It also provides great detail on how critical the upcountry was to the success of the Revolution.

<sup>3</sup> Lacy K. Ford makes a strong case for the importance of the Upcountry in the formation of the state's economic and political climate. He argues that the Upcountry's economy, defined by the "co-

Charleston planters, and the area ended up becoming the permanent home of such notable families as the Mavericks, Pickens, and Calhouns. The region's geographic diversity lent itself to more economic diversity than the Low Country, as well. Historian Scott Poole defined it as "the boisterous younger brother to the Low Country, coming of age in the nineteenth century cotton boom."<sup>4</sup>

The region developed a diverse economy, including many mills, local mercantiles, and, of course, agriculture. It is in this period that a free African American population coalesced into communities in the Upper Piedmont and plantations in the Lower Piedmont grew to include numerous slaves (Abbeville). This period witnessed the growth of the "resort industry," as wealthy planters from the Low Country came to the Greenville and Pendleton Districts to escape the heat and disease of the summers<sup>5</sup>—and often stayed, as in the case of the Calhoun, Broyles, and Maverick families. It also saw the growth of cotton as a cash crop at this time, although the majority of farmers continued to pursue subsistence farming first until the 1850s. The region was known for its grain mills and railroad developments. During the antebellum period, evangelical religion took hold as both a unifying and community-building force, and politics in the region looked much like that of other frontier districts—hard-fought, personal, and intensely democratic, marked by a very high voter turnout. The African American population grew in the Lower Piedmont district, but few plantations rivaled those of Low Country plantations, which contributed to the fact that many of the districts in the Low Country had over 80% African American populations.<sup>6</sup> Very often Upcountry slaveowners held just a few slaves, and so they tended to work alongside, worship with, and have more contact with their slaves. In many of the upper piedmont (Appalachian) districts, whites and African Americans belonged to the same churches and generally came in much closer contact with each other than did Low Country regions, with their much higher black populations.<sup>7</sup> While agriculture dominated the Upcountry, as it did throughout the South, distilleries, sawmills, and

---

existence of market involvement with the ethic of self-sufficiency," 74, drove agriculture to focus on subsistence first but with an eye toward cotton as a cash crop, encouraged the development of railroads, and fostered the growth of artisan manufacturing, distilling, and milling. He claimed that while the Upstate never was able to dominate politics in the state, the "intensely competitive" politics, 112, was open and incredibly democratic.

<sup>4</sup> Scott Poole, *Never Surrender*, 6.

<sup>5</sup> "Pendleton, SC, Town History,"

<http://www.smalltowngems.com/browsetowns/southcarolina/pendleton/pendletonsctownhistory.html>

<sup>6</sup> Ford, 32, 44, 46-7, 108-112, 123, 231, 253-9.

<sup>7</sup> W.J. Megginson's *African American Life in South Carolina's Upper Piedmont*, an exhaustive study of African Americans in the Anderson/Oconee/Pickens District of the region from the 1770s to 1900 reveals the fact that slaveholding populations in this region looked much more like the populations of other Appalachian regions of the South than it did the Low Country. (Comparison from selections in John Inscoe's edited collection *Appalachians and Race*, which deals at great length with slavery in Appalachia, but with no contributions on the districts that would become recognized as the seven "Appalachian" counties of South Carolina, as defined by the Appalachian Regional Council).

other ventures like railroads in the Upper Piedmont developed—and sometimes, as in the case of the Blue Ridge Railroad—stalled out.<sup>8</sup>

The Upcountry was equally known for its powerful politicians, who allied themselves with Low Country planter interests to rule yeoman farmers with an iron fist. Families were patriarchal first, with elite men employing paternalism to curry favor with white yeoman farmers and maintain their domination over the slaves. Underlying this paternalism lay a violent core. Edgefield District was notorious throughout the country from the colonial period through the twentieth century for its violence, which included numerous duels and shootings,<sup>9</sup> and indeed the lower Piedmont was known for secessionist politicians as early as the 1840s, according to Lacy Ford. Indeed, Senator Brooks was notorious for his attack on Charles Sumner on the Senate Floor, fueling the fire of the already divided nation in the 1850s. Although the region did not support secession as the low country did in the 1840s, by 1860, unlike most upcountry regions of different states, it did generally wholeheartedly support secession, partly because, according to Lacy Ford, residents of the region viewed secession as a defensive maneuver; they were supporting it to guard their “republican values” against the North, who sought to take away their independence.<sup>10</sup>

Sites that illustrate this boom period of the Upcountry include:

- *Ashtabula* is a historic plantation built in the 1830s that interprets both the lives of the many owners, the lives of planters living in Pendleton, and the lives of the slaves who lived there through guided tours, camps for children, and special living history events. Ashtabula was one of our test sites for the project.
- *Woodburn* is a historic plantation built in the 1830s by the Pinckney family and lived in by the Adger family throughout the antebellum and into the twentieth century. Woodburn also interprets the lives of its owners, as well as the slaves and tenant farmers who lived there, through guided tours and living history events. Woodburn was one of our test sites for the project.
- Built before the 1850s, the *Seay House* in Spartanburg interprets the lives of simple yeoman farmers who lived in log cabins like this one. The Seay House is unique in the fact that it focuses on the lives of farm women who inhabited this house through the twentieth century.
- Built in 1795, the *Price House* reflects the growth of industry, as well as infrastructure, in this newly settled region. The Price family ran a tavern, post office, inn, and store for residents and travelers to the region, which is now in Spartanburg County, using twenty-four slaves to assist with the labor. Today the Spartanburg Historic Association uses the house and its 200-year-

---

<sup>8</sup> Poole, 14-15; Ford, 231, 268-9; “Stumphouse Tunnel Historical Marker,” <http://www.hmdb.org/marker.asp?marker=15041>

<sup>9</sup>“Violence” section, “Historical Overview of Edgefield,” <http://www.historicedgefield.com/id3.html>

<sup>10</sup> Ford, 368, 372.

old buildings to tell the story of how free and enslaved residents of the area built up the economic and social structures of the region.

- The home of John C. Calhoun, and the site of his drafting of the infamous nullification treatise “Protest and Exposition” in the 1830s, *Fort Hill Plantation* represents the lifestyle of the slaveholding elite at the time. Guided tours detail the life of this Senator and Vice President, as well as the life of Thomas Green and Anna Calhoun Clemson, who deeded the land to Clemson University.
- No other geographic feature better illustrates the attempt of industries to penetrate the Upcountry than the *Stumphouse Tunnel*. The tunnel’s construction brought Irish immigrants to the Appalachian region of the upcountry in the 1840s. Hired to lay track for the Blue Ridge Railroad and blast their way through mountainous tunnels, these immigrants founded a Catholic church and brought diversity to the far upcountry. The tunnel was never completed, as the venture ran out of money just prior to the Civil War. Visitors can walk through the finished part of the tunnel in a Walhalla city park.

Although the Upcountry remained fairly isolated from most major Civil War fighting, it emerged into the post-war era a place forever changed. As Scott Poole argues, the Upcountry elites’ desire to “stabilize the ownership of productive property rightly viewing mastery over land as a prerequisite to for mastery in social relations,” led to its leadership in numerous Lost Cause memorials and tributes.<sup>11</sup> The era saw the emergence of race-baiter “Pitchfork” Benjamin Tillman of Edgefield, who sought favor from poor whites by fighting to support agricultural college to be built in the upstate, and by race-baiting and supporting segregation in all of its forms.<sup>12</sup> In essence, Scott Poole argues that the Upcountry became more conservative as elites tried to uphold “premodern values” based on the control of dependents, no matter who they were. As elites sought industrialization and modernization on the one hand, they fought to retain an antebellum social order, a situation that led Poole to define the area as “a theater of contrast and change.”<sup>13</sup>

Jim Crow laws had the same effect in the upstate as they did elsewhere. The African American community responded with activism. Vernon Burton and W. J. Megginson note that religion, education, and family were the mainstays of the African American community at this time.<sup>14</sup> African American business districts

---

<sup>11</sup> Poole, 58.

<sup>12</sup> Stephen Kantrowitz’s *Ben Tillman and the Reconstruction of White Supremacy* provides a detailed account of the populist’s political agenda and popularity with the white lower class. Poole argues that Tillman was representative of the contested nature of the upstate at the time—while he was for progress, he squarely aligned himself with agricultural interests, 158-9.

<sup>13</sup> Poole, 22, 6.

<sup>14</sup> Vernon Burton’s *In My Father’s House Are Many Mansions* studies black and white families before and after the Civil War in Edgefield, SC. In this extensive quantitative analysis of households, he argues that family cohesion and religion were critical to the survival of black families—and indeed, black families were put under pressure especially when African American men could not find work in the region that would support their families. He claims that white and black families considered a

thrived in both Pendleton and Anderson, and newly freed people moved to the new town of Seneca and the burgeoning cities of Greenville and Spartanburg to seek out new opportunities, although, as Megginson points out, most African Americans remained in the agricultural and domestic service fields. Still, African American education and fraternal organizations flourished in the period, as black residents of the Upcountry sought to claim and retain their rights in the period.

Although the region remained primarily rural, this period witnessed intense industrialization in the Upcountry, and historians Melissa Walker and Bryant Simon chronicle the movement of white families into mill towns and villages. The “New South” came quickly to the Upstate, as investors from the North coupled with southern speculators to bring industry to the area. Although some men worked at the mills to save their family land, often white women enjoyed the more modern conveniences of the mill homes and enjoyed the lack of isolation they experienced in towns. As historians commonly claim, early twentieth-century mills of the South were known for their long hours, lack of unionization, and use of child labor. Still, for many whites, mill life offered opportunities that were lacking in the agricultural sector, as Walker notes.<sup>15</sup>

Select sites that cover the Reconstruction and Industrialization Era include:

- The *Lunney Museum* of Seneca is a Queen-Anne style bungalow that typifies the New South lifestyle. Owned by an upper-middle class family, the Lunney Museum retains artifacts and has architecture that reveals the gracious living typified by men who made their money in trade, rather than on the land. This site has already been developed.
- *Bertha Strickland Cultural Museum* and the *Seneca Institute* highlight African American history in Seneca, and reflect the growth of a prosperous and educated middle-class that took advantage of opportunities in a New South town.
- *Silver Spring Baptist Church* in Pendleton is an African American church, built in 1874, that educated many children in the community, including Jane Edna Hunter, who became part of the Progressive Movement in Cleveland, founding the Phillis Wheatley Association there and ultimately building it into a national institution.
- Financed by a Charleston Native, the *Newry Mill* opened its doors in 1894 as the first mill in Oconee County. The mill functioned until 1975, and many residents left to find work elsewhere. The village typifies the traditional mill town at that time—isolated from the rest of Oconee County, its 118

---

patriarchal family to be the basis of an orderly society. Megginson claims that African Americans fought to create familial, socio-religious, and educational institutions to ameliorate the indignities of racism in the region.

<sup>15</sup> While Melissa Walker’s *Southern Farmers and their Stories* and *All We Knew Was to Farm* examines the lives of farmers in the South, she discusses the importance of textile mill money and towns to rural white families. Jacquelyn Dowd Hall and Robert Korstad’s *Like a Family* details the culture and society of southern mill towns. Simon’s *A Fabric of Defeat* focuses mainly on the politics of race among white workers in the mills from 1910 to 1948.

structures sit on 250 acres of land. Historic homes range from small village bungalows to churches and mill managers' homes.<sup>16</sup>

The Upcountry faced a series of upheavals from the 1930s through the end of the twentieth century. Like the rest of the South, it was beset by agricultural and financial crises in the 1930s, civil rights unrest from the 1940s through the 1960s, and industrial collapse in the late twentieth century as a result of outsourcing. Furthermore, the flooding of the Seneca River dislocated traditional Upper Piedmont Communities as it created a new tourism venue with Lakes Hartwell and Keowee. In the 1930s and 1940s, Upper Piedmont residents showed that they were not as docile as "New South" boosters had claimed, as historian G.C. Waldrep III notes in his study of Spartanburg County mills. Bryant Simon notes that white men's growing concern over their own position and their determination to remain dominant economic actors led many to support waves of unionization efforts as early as the 1920s. By the 1930s, as he notes, workers' actions showed increasing activism as they fought for better conditions, wages, and the right to unionize. Indeed, Chiquola Mills in Honea Path became the symbol of employer brutality in the Great Textile Strike of 1934 as seven strikers were gunned down as they picketed the mill. According to Simon, labor unrest continued throughout the 1940s. Worker unrest, along with new labor laws instituted in the 1930s, may have ameliorated the worst factory conditions, but nothing could stop the movement of textile mills overseas in the late twentieth century as employers looked for cheap labor and less stringent environmental laws. Today, shells of mills dot the Upcountry landscape—some have been converted into loft apartments and dining/entertainment areas, as in Greenville, but most have been torn down or left to deteriorate.

African Americans emerged from South Carolina's Depression more activist, with stronger institutions that focused on civil rights, like the NAACP. After World War II, the lynching of Willie Earle, which began in Easley and ended in Greenville, led to what historian William Grady termed the "largest lynching trial in southern—perhaps in American—history" in 1947.<sup>17</sup> Earle had been accused of beating a cab driver to death, and a mob of whites brutally beat and shot him before throwing him over a bridge. While most whites merely assumed that Earle was guilty, and the mob participants were acquitted because they focused on Earle's presumed guilt, the lynching drew attention from both black and white groups. African Americans protested the lynching and argued for the "civil right not to be lynched" while whites focused on the danger of letting mobs run amok without instilling law and order. Strom Thurmond, then governor, offered to send reinforcements to local sheriff units, and four years later the state passed an anti-lynching law. Still, the tide had turned, and African Americans grew more vocal in their demands for equality. By 1954, a Low Country South Carolina court case subsumed under *Brown v. Board of Education* began the process of reshaping the racial parameters of South Carolina and the rest of the country. Civil Rights actions continued in the Upstate. In 1960,

---

<sup>16</sup><http://www.sciway.net/sc-photos/oconee-county/newry-mill.html>

<sup>17</sup> William Grady, "The Civil Right to Not Be Lynched," 94.

the “Greenville 8,” a group of college students which included Jesse Jackson, successfully integrated the main branch of the Greenville Public Library system, just months after the famous Greensboro sit-in.<sup>18</sup> Three years later, Clemson College reluctantly opened its doors to Harvey Gantt, thus becoming the first integrated college in the state. Although the college experienced no violence the day Gantt matriculated, unlike events in Mississippi and Alabama, Gantt’s experience at Clemson somewhat belied the storied tale of “Integration with Dignity,” according to Vernon Burton.<sup>19</sup> Still, the tradition of separate schools was broken, at least at the postsecondary level. It would take until the 1970s for Upcountry primary and secondary schools to desegregate.

Agriculture remained a mainstay of the Upcountry in this period, although it was beset by falling prices and soil exhaustion, as well as the vagaries of the national economy. According to Melissa Walker, farmers experienced the Depression very differently, depending on their race and class. As in other places, African American farmers failed to receive the same benefits from New Deal programs as their white counterparts, and white landowners made out the best. As in the rest of the South, tenant farmers and sharecroppers were often pushed off the land as landowners took federal money to mechanize production and keep fields fallow. Walker’s study of women in the Upper South during this period sheds light on the ways in which families negotiated the change, handling the demands of federal government officials and attempting to retain their families’ autonomy in the face of massive economic dislocation.<sup>20</sup> Further change came to farmers, rich and poor alike, when Duke Energy flooded the Seneca River to make way for hydroelectric and nuclear power in the mountainous regions of Oconee and Pickens counties in the 1960s and 1970s. Scholars have yet to delve into the dislocation experienced by longtime residents, but there is work to be done to uncover the feelings of those who were displaced. Entire villages were removed to make way for “progress,” and often unwillingly. The Upcountry is known for its peach production, which began in the nineteenth century but took off in the 1930s and continues to this day—in fact, South Carolina is the number two producer of peaches nationally, and grows more peaches than Georgia.<sup>21</sup> Today, the Upcountry is home to many small family farms that produce for the burgeoning “locavore” movement, so the tradition of farming that began in the colonial period is continuing to this day.

Sites that interpret this era of history include:

---

<sup>18</sup> <http://www.wyff4.com/-Greenville-8-Together-50-Years-After-Segregation/-/9324882/6161420/-/bw6y7x/-/index.html>

<sup>19</sup> Burton, Orville Vernon. "Dining with Harvey Gantt."

<sup>20</sup> Several of Melissa Walker’s books focus on farmers in the South, but she includes a great deal of information about upcountry farming in her work. Focusing on oral history, she finds that communities were intrinsically important to the survival of family farms—although “community” held different meanings for famers of different races and classes.

<sup>21</sup> “A Historical Retrospective of South Carolina’s Peach Industry,” <http://agriculture.sc.gov/content.aspx?ContentID=756>.

- Although the *Chiqoula Mill* shut down in the 1990s and is currently abandoned and deteriorating, there is a historic marker there that could serve as the focus of a virtual site, much like what we did with the Seneca Institute site.
- A plaque commemorating “Integration with Dignity” at *Clemson University* could serve as the focus here for a story about Harvey Gantt and the integration of South Carolina’s universities.
- Underneath *Lake Jocassee* lie some of the town structures that were flooded to make way for hydroelectric power in the 1970s, including a vacation lodge built in the 1920s.<sup>22</sup>
- Built in 1938 by the WPA, *Greenville High School* still stands as a testament to the ways in which federal money helped in the Upcountry.

In addition to those sites listed above, there are many larger history museums in the region that seek to tell the entire story of their counties from the colonial era to the present.

- The Upcountry History Museum
- Anderson County Museum
- Oconee Heritage Center
- Pickens County Museum/Hagood Mill
- Aiken Museum
- Spartanburg Regional History Museum

*The Palmetto History Project* has already received commitments from many of these heritage tourism sites to collaborate in developing appropriate content for use in the proposed mobile apps and we will work with other sites as time, money, and interest allow to ensure that our project covers the history of the Upcountry as holistically as possible. Indeed, our intent is that *The Palmetto History Project* will not simply discuss particular sites; instead, it places all sites in a broader southern, and national historical narrative context. As such, the larger goal of this project is to center the Upcountry’s citizens as important historical actors in national, Appalachian, Southern, and African American history and facilitates conversations about the meanings of local sites and historical identity. It encourages tourists to leave the “beaten path” and to discover cultural sites about which they would otherwise not know, introducing them to an important region that deserves more national attention.

We contend that these historic sites are critical to understanding the narratives of the upcountry, as they interpret history that has not yet been fully plumbed by scholarly literature. In this sense, they are producing a synthetic narrative that historians have not. We cannot fully grasp both the complexity and relevance of the region’s history without exploring these sites.

---

<sup>22</sup> <http://awriterinthewry.wordpress.com/2005/05/06/underneath-the-lakes-tranquil-waters-lie-jocassee-memories/>

Public historians were among the first in our field to understand the significant impact digitization would have on the way academics research and museum patrons experience history. Museum scholars, including Stephen Weil, and Nina Simon have argued that to survive, museums, archives, and historic sites must engage with the public and move from traditional didactic educational spaces to more forum-based interactive sites for community conversation, education, and entertainment.<sup>23</sup> From digital archives to technology in exhibits, public historians have been working on various ways to increase their ability to connect with museumgoers. Digital collections have become invaluable tools for historians researching from offsite, and virtual exhibits and even entire virtual museums have sprung up to inform and entertain users that may never set foot in a “bricks and mortar” site. Museums are attempting to embrace new exhibit technologies that enable users to interface with the curators and other visitors, including downloadable podcasts that serve as tour audio throughout exhibits, video talkback recording areas, computer games, and even digital tags within museums that can be scanned by a mobile device to provide more information about particular subjects. Temporary exhibits are becoming archived as virtual ones so that visitors can always go back and see the images and artifacts from a certain collection.

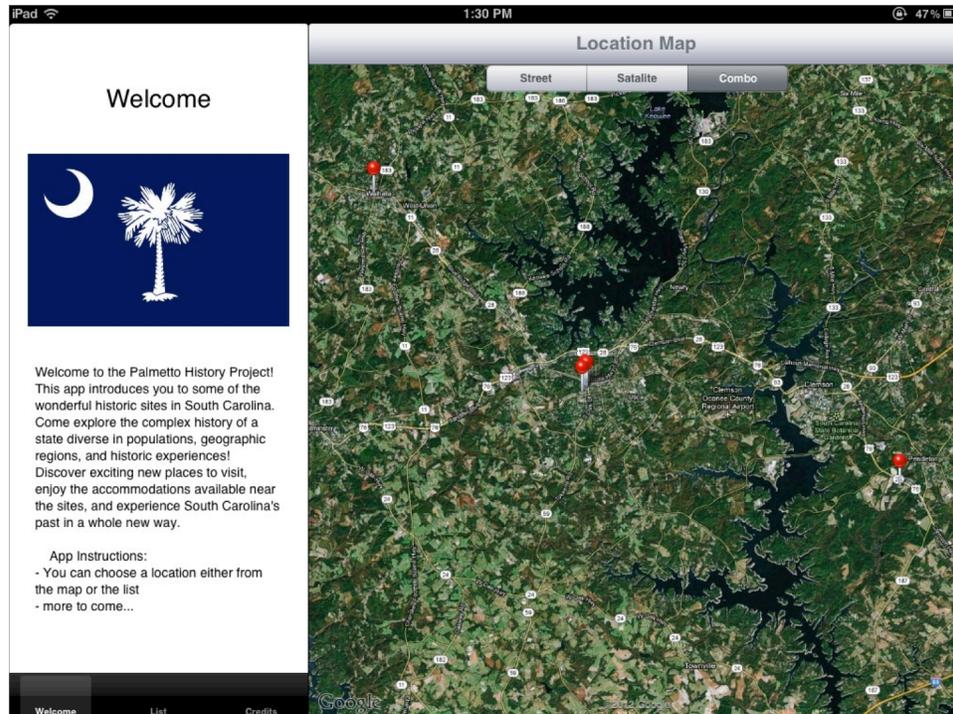
*The Palmetto History Project* uses the best of new technologies in a way that strikes a balance between excellent and compelling virtual media and site-based content that will drive users to visit the sites featured. Videos, accompanying text, and links to pictures and archival material will not only enhance a visitor’s experience, but also will encourage the potential visitor to visit in the first place. Local yelp reviews are linked to the specific sites, which provide a community forum for visitors to review sites and surrounding hotel and dining facilities. In so doing, we will help these sites become, in Nina Simon’s analysis, “participatory museums,” ones that will serve their communities, educate visitors, and facilitate ongoing conversations about Upstate South Carolina history<sup>24</sup>

### **Project Format and Participant Experience**

*The Palmetto History Project* has brought together historians, students, technology experts, and historic site directors to create and populate a computational infrastructure that allows users to explore historical sites in innovative ways whether they are at home sitting in front of a computer or physically at the site. Key features of the project include a dedicated mobile application (in this prototype phase, developed for Apple’s iOS devices, but in the future also made available on Android and Windows Mobile 7) that provides the following functionality:

---

<sup>23</sup> Many museum studies scholars have pointed to the need to reach out to potential and on-site visitors in different ways. Nina Simon contends that a “scaffolded experience,” one that allows for visitor participation within a structured environment works best (*The Participatory Museum*) 13. Our app provides the scaffolding, as the content is created by us, but it is participatory in that users can choose which sites to “visit,” as well as which links to experience on each site. Other scholars who focus on how to include museumgoers in participatory experiences include Stephen Weil, *Making Museums Matter*; and the contributors to *Letting Go? Sharing Authority in a User-Generated World*, eds. Bill Adair et. al.



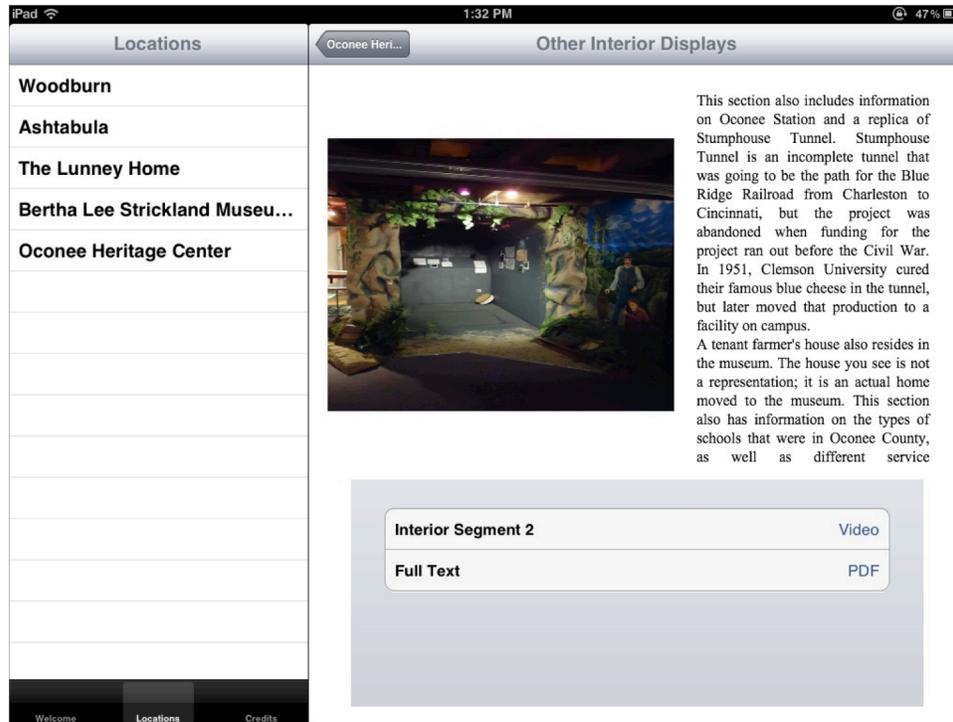
**Figure 2: Splash Screen for the Prototype *Palmetto History Project* iPad App**

- Detailed historical narratives describing a site's historical significance
- Access to relevant archival materials, such as historic photographs
- Short video documentaries addressing the site's history and its significance in the larger southern historical narrative
- Mapping through Google maps that enables users to view nearby historic sites of interest
- Information about local lodging options and restaurants

These experiences are entirely customizable—users can choose the ways in which they engage with the app. Material is chunked for each site, so that users can decide which segments interest them most and view them as they please. Text and video are available, which enables differently abled users to explore the sites virtually. Links to the historic sites' websites provide more information to potential tourists and educators looking to plan class trips.

### **Project Resources**

*The Palmetto History Project* mobile app takes advantage of advances in software and hardware to help users discover sites of historic interest in their vicinity. One key feature will be the ability of the app to alert users to a nearby site based on user-defined criteria such as vicinity, and type of site. Each site within our database will have its own dedicated page from where users will be able to access content such as digitally archived resources for researchers, virtual tours, and interviews with curators, interpreters, and historians. Although this mobile app has initially been developed for Apple's iOS system, we are fully committed to developing versions for



**Figure 3: Sample Page From the Prototype iPad App**

other devices—such as Android and Windows Mobile 7—in later iterations of the project and as funding allows. Even so, non-iOS users will be able to access the majority of our content from their devices using our HTML 5-compliant website.

Content for this project will come from three principal sources. First, we work closely with our partner historic sites to identify existing materials and create narratives that sensitively and accurately depict the wider significance of each locale. Second, students working with PI Megan Shockley as part of her Public History Emphasis Area at Clemson University will visit each site and take photographs, create narratives, and interview staff site to weave site narratives into broader research papers that serve as the script for the student-produced documentaries. Students working with co-PI Roy Pargas are coding the content for the app. Finally, our website, which is under construction, will offer the ability for users to contribute their own experiences and photographs of each site. This final category of content will be moderated by the project team. Metadata will be repurposed from existing databases and finding aids held by the partners and will be enhanced as needed for consistency. This “talkback” portion of the project will encourage conversation between a diverse museum going public and site directors, as well as the students involved in the project. In this way, the digital media can facilitate communication across geographic, cultural, and generational boundaries, and the audience is limited only by its ability to access a website.

Clemson University is providing the server space to host the site, and we have the support of museum directors who have agreed to participate in the project. Thus far, the site directors have given the content developers free and unrestricted access to site scripts, images, and other relevant material.

## **Project History**

The genesis of *The Palmetto History Project* came about as a result of Principal Investigator Megan Shockley's extensive work with heritage tourism sites in Upstate South Carolina. Working with staff at Clemson University's CyberInstitute and local sites, she identified a significant need for a program that would simultaneously engage humanities students in digitization/alternate media projects and highlight South Carolina as important in the larger regional/national historical narratives scholars currently produce. In evaluating existing heritage tourism apps, we determined that we wanted to create a larger-scale program than that, with many more types of content and a scholarly component that would assist those who were not casual travelers through the area to learn more about South Carolina and its history. Although still in the relatively early stages, *The Palmetto History Project* has already garnered support from both Clemson University and museum sites in the upstate. The Clemson CyberInstitute has awarded the project \$5000 in initial seed funds, and the Clemson University Service Alliance awarded \$1500 in funding to enable the PIs to create the initial app. The test app received the generous support of three boards of directors in historic sites, which resulted in the creation of five different historic site modules.

## **Audience, Marketing, and Promotion**

The intended audience includes: teachers planning field trips; tourists interested in planning a visit to historical sites; casual tourists looking for activities in their targeted vacation area; residents of the Upcountry who have not experienced the wealth of historic sites in their own area; casual users of the website and app, who may never visit the area but can learn about the importance of Upstate history as "virtual tourists." We plan to market this app in several ways:

- Discuss the app in a "Your Day" segment, a widely-heard state-wide NPR lunchtime program.
- Launch the app formally by notifying the South Carolina Federation of Museums (<http://www.southcarolinamuseums.org/>); the South Carolina African American Heritage Commission (<http://shpo.sc.gov/res/Pages/SCAAHC.aspx>); the State Historic Preservation Office (<http://shpo.sc.gov/Pages/default.aspx>); and The Consortium for Upstate South Carolina History (<http://cusch.wordpress.com/2012/03/26/upcoming-events-at-the-upcountry-history-museum/>). We believe that these organizations will help us to get the word out to their significant membership lists with announcements in newsletters and on listservs.
- Notify the major CVBs in the area with a request to provide information on the app and with a link to the app store. These CVBs include: The Pendleton District Commission (<http://www.pendletondistrict.org/>); Mountain Lakes Convention and Visitors Bureau (<http://www.scmountainlakes.com/>); "Upcountry South Carolina" (<http://www.upcountrysc.com/>); the South

Carolina National Heritage Corridor ([www.sc-heritagecorridor.org](http://www.sc-heritagecorridor.org)); and the numerous local county CVBs in the upcountry.

We recognize that although this digital project can reach out via the internet to thousands of potential users, we cannot reach casual travelers this way. We plan to produce a pamphlet highlighting the sites, the project, and the app and companion websites that can be stocked at the CVBs, local museums, and rest stops in South Carolina.

### **Evaluation, Final Product, and Dissemination**

There are several ways in which we can gauge the success of this project. The first will be to have the sites we develop compare their past-year visitor data with the post-project numbers. We can also ask the sites to do a quick survey asking visitors how they heard about the location. In addition, we can track hits on the apps and the companion website, and we can track the user reviews that will inevitably arise with the usage of the apps. We would obviously like to see how much traffic we could drive to an area with this app, but just as important is educating the public about South Carolina history. Tracking the outcomes using these separate but related data sets will enable us to see where we are having the most impact.

The primary product of this grant will be an expanded website and mobile app with documentary movies that provide overviews of historic sites and links to amenities in each area. In addition, we will generate a white paper and final report detailing the technical and historic work conducted through the grant period. Finally, this project will result in further submissions to funding agencies, including the South Carolina Heritage Corridor, the South Carolina Humanities Council, and the South Carolina Tourism Board.

### **Organizational Profiles**

Clemson University was founded in 1889, a legacy of Thomas Green Clemson, who willed his Fort Hill plantation home, its surrounding farmlands and forest, and other property to the state of South Carolina to establish a technical and scientific institution for South Carolina. Clemson opened its doors to 446 students as a military college in 1893. Today, Clemson is classified by the Carnegie Foundation as a Doctoral/Research University-Extensive, a category comprising less than 4 percent of all universities in America. Students can choose from more than 70 undergraduate and 100 graduate degree programs in five colleges. As the state's land-grant university, Clemson reaches out to citizens, communities, and businesses all over South Carolina through county-based Cooperative Extension offices, five off-campus Research and Education Centers, and critical regulatory responsibilities for plant and animal health. Clemson University is governed by a board of thirteen members, including six elected by the State General Assembly and seven appointed self-perpetuating life members. Clemson University operates as a unit of the State of South Carolina (the primary government) as a state assisted institution of higher education. The University's trustees approved a budget of \$859.8M for FY2012-13.

The College of Architecture, Arts and Humanities is organized into three schools, offering a wide range of academic and professional programs through its

ten academic departments—from the visual and performing arts to the design and building disciplines to the humanities. The University Special Collections Division holds numerous archives relating to the history of the Upcountry, including the John Caldwell Calhoun Papers, the Thomas Green Clemson Papers, and the Strom Thurmond Collection.

The Spartanburg County Historical Association was founded in 1957 to preserve the history of the region. It owns the Spartanburg Regional History Museum, Walnut Grove, the Price House, and the Seay House. It hosts numerous programs annually, including historic re-enactments, summer camps, and other special events with the help of over 200 volunteers. With a budget of \$340,000 and a full-time staff of three with 20 regular volunteers, the Association maintains and runs these sites for the 15,000 visitors it sees each year. It also houses a manuscript collection relevant to the region and has helped to produce numerous historical monographs about the area.

### **Project Team**

*The Palmetto History Project* will be led by Professors Megan Shockley and Roy Pargas of Clemson University. Shockley's research focus is African American history and women's history. She is the author of two books and one forthcoming book, as well as several articles. She has administrative experience as secretary of the Southern Association for Women in History, a position in which she served for five years.

Dr. Roy Pargas teaches in the Human Centered Computing Division of Clemson's School of Computing. His research expertise is in *teaching with technology*. Over the past twelve years, he has designed and implemented software applications (apps) used in teaching college chemistry, calculus, engineering, nursing, history, biological sciences and computer science. His work has been supported by over \$1.5 million in grants from the National Science Foundation, Microsoft, Hewlett-Packard, the Dreyfus Foundation, the U.S. Department of Education and Clemson University. He has written over 60 papers in computer science journals and conference proceedings as well as chapters in two books.

This project will be supervised by a Board of Scholars who are intimately connected with the Upcountry and its history. The Board will be chaired by Vernon Burton, Director of the Clemson CyberInstitute, and a Professor of History and Computer Science at Clemson University. Burton was the founding Director of the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois, where he is emeritus University Distinguished Teacher/Scholar, University Scholar, and Professor of History, African American Studies, and Sociology. Burton has authored or edited sixteen books and more than two hundred articles.

In addition to Burton, a number of notable scholars who have explored the region in their work have agreed to serve on the advisory board, including: Lacy K. Ford, University of South Carolina; Carmen Harris, Associate Professor of History at University of South Carolina-Upstate; Becky Slayton, Director of the Spartanburg Historical Association; and Melissa Walker, George Dean Johnson Chair of History, Converse College.

## Work Plan

Work on the project, which will be conducted between April 2013 and March 2014, will fall under two themes: historical content creation and technical development. First, we will bring together a team of prominent scholars whose work focuses on the Upcountry in order to prioritize site development. Second, we will expand the historical content that will populate the site. PI Shockley will supervise site surveys and research that will be conducted by student workers. The PIs will also work closely with library and information science staff at Clemson University to ensure that all generated content is properly created in a standards-based format that will allow it to be easily ingested by the content management system. This work continues what was begun in Spring 2011. Work on the project will be conducted according to the following plan:

	A	M	J	J	A	S	O	N	D	J	F	M
<b>Advisory Board</b>												
Initial Meeting at Clemson		✓										
Review of Prototype Apps								✓	✓			
Second Meeting at Clemson										✓		
<b>Content Creation</b>												
Identification of Heritage Tourism Sites to be Added to the App	✓	✓										
Creation of Historical Content in Partnership with Heritage Tourism Sites		✓	✓	✓								
Populate Mobile App Database with Content				✓	✓	✓	✓					
<b>App Development</b>	✓	✓	✓									
Identification of Standards and Underlying Infrastructure	✓	✓										
Refinement of Existing Prototype App	✓	✓	✓	✓	✓	✓						
Development of Android Version of App						✓	✓	✓	✓			
Beta Testing of Apps							✓	✓	✓			
Continued Refinement of Mobile Apps									✓	✓	✓	✓
<b>Final Report</b>											✓	✓

We have identified four sites to populate the next phase of the Palmetto Heritage Project. Walnut Grove Plantation and Price House are owned by the Spartanburg Historical Association and interpret the colonial period through the mid-nineteenth century. One, Newry Mill Village Historic District has no historical society or organization with which to work, although it is on the National Historic Register. The defunct historical society placed records at the Oconee Heritage Center, with whom we have a strong working relationship. Additional information about Newry Village is available from the National Trust application, manuscripts in Clemson University Special Collections, and numerous dissertations produced by Clemson University students and available in the library. This site is a National Historic Trust site. The fourth site, Jocassee Village, will be a completely virtual site. The village is under water, and is a popular dive site located in a state park. As with the Seneca Institute site, students will compile information from various archival sources, including newspapers available at the Oconee County Public Library and Clemson

University, and oral histories and other archival collections at Clemson University Special Collections.

**Fundraising Plan**

Outside of applying for grants from the Watson Brown Foundation, the South Carolina Council on the Humanities, the University Research Grant, and other grants from various funding agencies, we have no plans to raise funds for this project at this time.

## **Work Samples for Digital Components**

Examples of video components already produced can be viewed at:

- <http://www.youtube.com/user/museumprofessor?feature=mhee>

To access a beta version of the prototype iPad app (requires iOS v 5.0 or higher), please visit:

- <http://testflightapp.com/dashboard/team/accept/24414c051a81ec7204b245278ed5d5ed0c70392d/>

## Bibliography

- Adair, Bill, Benjamin Filene, and Laura Koloski. Eds. *Letting Go? Sharing Authority in a User-Generated World*. Philadelphia: Pew Center for Arts and Heritage, 2011.
- Burton, Orville Vernon. "Dining with Harvey Gantt: Myth and Realities of 'Integration with Dignity.'" In *Matthew J. Perry: the Man, His Times, and His Legacy*, edited by W. Lewis Burke, Jr., and Belinda F. Gergel. Columbia: University of South Carolina Press, 2004.
- Burton, Orville Vernon. *In My Father's House are Many Mansions: Family and Community in Edgefield, SC*. Chapel Hill: University of North Carolina Press, 1987.
- Edgar, Walter. *Partisans and Redcoats: The Southern Conflict that Turned the Tide of the American Revolution*. New York: William Morrow, 2003.
- Ford, Lacy K. *The Origins of Southern Radicalism: The South Carolina Upcountry, 1800-1860*. New York: Oxford University Press, 1991.
- Grady, William. "The Civil Right Not to Be Lynched: State Law, Government, and Citizen Response to the Killing of Willie Earle." In *Toward the Meeting of the Waters: Currents in the Civil Rights Movement of South Carolina During the Twentieth Century*. Eds. O. Vernon Burton and Winifred Moore. Columbia: University of South Carolina Press, 2008. 93-119.
- Hall, Jacquelyn Dowd and Robert Korstad. *Like a Family: The Making of a Southern Cotton Mill World*. Chapel Hill: University of North Carolina Press, 2000.
- Hatley, Tom. *The Dividing Paths: Cherokees and South Carolinians Through the Era of Revolution*. New York: Oxford University Press, 1995.
- Inscoc, John. *Appalachians and Race: The Mountain South from Slavery to Segregation*. Lexington: University Press of Kentucky, 2000.
- Kantrowitz, Stephen. *Ben Tillman and the Reconstruction of White Supremacy*. Chapel Hill: University of North Carolina Press, 2000.
- Megginson, W.J. *African American Life in South Carolina's Upper Piedmont*. Columbia: University of South Carolina Press, 2006.
- Poole, Scott. *Never Surrender: Confederate Memory and Conservatism in the South Carolina Upcountry*. Athens: University of Georgia Press, 2004.
- Simon, Bryant. *A Fabric of Defeat: The Politics of South Carolina Millhands, 1910-1948*. Chapel Hill: University of North Carolina Press, 1998.
- Simon, Nina. *The Participatory Museum*. Museum 2.0, 2010.
- Walker, Melissa. *All We Knew Was to Farm: Rural Women in the Upcountry South, 1919-1941*. Baltimore: Johns Hopkins University Press, 2002.
- \_\_\_\_\_. *Southern Farmers and their Stories: Memory and Meaning in Oral History*. Lexington: The University Press of Kentucky, 2009.
- Weil, Stephen. *Making Museums Matter*. Washington, D.C.: Smithsonian Books, 2002.



**Center for Digital Humanities**  
University of South Carolina  
Department of English  
Humanities Office Building  
Columbia, South Carolina 29208

David Lee Miller, Director  
dmill1951@gmail.com

September 14, 2010

Re: Digital Humanities Start Up Grant (Level II): The Sapeos Project:  
Transparency in Multi-image Collation, Analysis, and Representation

On behalf of the University of South Carolina, I am writing today to request a change in Primary Investigator for the Digital Humanities Start Up Grant (Level II): The Sapeos Project: Transparency in Multi-image Collation, Analysis, and Representation. Awarded in August 2009, the Sapeos Project began its tenure of the direction of our Associate Director of the Center for Digital Humanities, Randall Cream. Dr. Cream worked in partnership with myself and Dr. Song Wang of the Department of Computer Science and Engineering while supervising the work of three graduate students on the construction of the Saphaeos image collation tool.

In May of this year, Dr. Cream left the University of South Carolina and I began to supervise the project work in conjunction with Dr. Wang. Dr. Jennifer Guiliano, the new Associate Director of the Center for Digital Humanities and Research Assistant Professor in History took over all roles and responsibilities assigned to Dr. Cream as part of this grant and will complete all tasks as projected including finalizing project work by December 30<sup>th</sup> of this year and facilitating a final white paper.

Given the commitment to continuing the work of the project, I write today to request changing the project PI from Dr. Cream to Dr. Guiliano. I do not anticipate any delays or difficulties with the change in Primary Investigators. Attached to this request is Dr. Guiliano's curriculum vitae. Please do not hesitate to let me know if you have any questions.

Sincerely,

A handwritten signature in black ink, appearing to read "David Lee Miller".

David Lee Miller  
Director, Center for Digital Humanities  
University of South Carolina

A handwritten signature in black ink, appearing to read "Jennifer Guiliano".

Jennifer Guiliano  
Associate Director, CDH  
University of South Carolina

**Citation:**

Brown, Travis and Jennifer Guiliano. "Topic Modeling for Humanities Workshop" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities  
Travis Brown  
Jennifer Guiliano  
Kirsten Keister  
Amanda Visconti

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## Topic Modeling For Humanities Research: Level I

**Enhancing the humanities through innovation:** Topic modeling is a statistical technique that attempts to infer the structure of a text corpus on the basis of minimal critical assumptions. One widely used topic model is Latent Dirichlet Allocation, which employs the following hypothetical story about how documents are created: we assume that each document is made up of a random mixture of categories, which we'll call *topics*, and that each of these topics is defined by its preference for some words over others. Given this story, we would create a new document by first picking a mixture of topics and then a set of words, by repeatedly choosing at random first one of the document's topics and then a word based on the preferences of that topic. This obviously isn't how documents are actually created, but these simple assumptions allow the topic model to work in reverse, learning topics and their word preferences by assuming that this story explains the distribution of words in a given collection of documents. A sub-area within the larger field of natural language processing, topic modeling in the digital humanities is frequently framed within a "distant reading" paradigm, drawing upon the 2005 work of Franco Moretti in *Graphs, Maps, Trees*. Moretti argues that the intersections of history, geography, and evolutionary theory create the potential for quantitative data modeling. Categorized as network theory, Moretti's approach to topic modeling utilizes aggregated data to explore macro level trends, themes, exchanges, and patterns. Yet, humanities scholars often need to focus simultaneously on the macro (distant/many texts) and the micro (close/individual texts). As a result, visualization between the "distant" aspect of the text's high-level attributes — its "topics" — and the "close" aspects the text — its individual words — are crucial. Latent Dirichlet Allocation (LDA) topic modeling provides a way to explore "distant" and "close" simultaneously. "LDA is a three-level hierarchical Bayesian model in which each item of a collection is modeled as a finite mixture over an underlying set of topics," wrote David Blei, Andrew Ng, and Michael Jordan in their 2003 article defining this field. "Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities." LDA offers an "unsupervised" topic modeling approach, in which no knowledge of the content of the text is really needed — the algorithm simply cranks away at whatever text corpus it is working on, and discovers topics from it — and a "supervised" approach where scholars "train" the algorithm by making use of domain knowledge. For example in a supervised LDA approach to Civil War newspapers, related pieces of knowledge coming from contemporaneous sources external to a corpus are used as additional data source. Casualty rate data for each week and the Consumer Price Index for each month allow the algorithm to potentially discover more "meaningful" topics if it has a way to make use of feedback regarding how well the topics discovered by it are associated with one of these parameters of interest. Thus, the algorithm can be biased into discovering topics that pertain more directly to the Civil War and its effects. Even this "supervised" version of LDA is not supervised in the sense generally used in machine learning, in which it indicates that the learner has been trained on labeled data and is attempting to assign these same labels to new data. As an unsupervised machine learning technique, LDA topic modeling does not require any form of expensive human annotation, which is often unavailable for specific literary or historical domains and corpora, and it has the additional benefit of handling transcription errors more robustly than many other natural language processing methods. Topic modeling aspires to discover global properties and qualities of the text, while at the same time connecting those global, macro-level qualities to micro-level detail, and is therefore likely to appeal to humanities scholars in a way that purely distant approaches do not. It is an approach that not only answers pre-existing research questions but also generates new questions. Despite—or perhaps because of—the relatively widespread use of topic modeling for text analysis in the digital humanities, it is common to find examples of misapplication and misinterpretation of the technique and its output. There are a number of reasons for this: existing software packages generally have a significant learning curve, most humanists do not have a clear understanding of the underlying statistical methods and models, and there is still limited documentation of best practices for the

application of the methods to humanities research questions. As a result, the most promising work in topic modeling is being done not by humanists exploring literary or historical corpora but instead by scholars working in natural language processing and information retrieval. These scholars, even as they have generated promising new avenues of research, have recognized topic modeling as “something of a fad” and suggested that more attention should be paid to the wider context of latent variable modeling approaches. This one-day workshop will facilitate a unique opportunity for cross-fertilization, information exchange, and collaboration between and among humanities scholars and researchers in natural language processing on the subject of topic modeling applications and methods. Recent work in natural language processing has particular relevance for research questions in the humanities, including a range of extensions of the basic LDA model that incorporate time and geography. Our intent is to begin to repair the divide between humanities using topic modeling approaches/software and those developing and utilizing them in computer science and natural language processing. Our primary goals will be: 1) greater familiarity with the interpretation and vocabulary of LDA topic modeling (and other latent variable modeling methods) for scholars in the humanities; 2) a deeper understanding of literary and historical corpora and their role as data within topic modeling; and 3) increased involvement in articulating fundamental research questions for researchers developing the models and methods (as well as the software implementations).

**Environmental scan:** Existing work in topic modeling and the digital humanities follows two major LDA approaches: synchronic, where the unit of analysis is not time bound, and diachronic, where the unit of analysis includes a measurement of time. Examples of synchronic work include Jeff Drouin’s exploration of Proust and Brown’s own work on Byron’s narrative poem *Don Juan* and Jane Austen’s *Emma* while examples of diachronic work include David Newman and Sharon Block’s work on the *Pennsylvania Gazette*, Cameron Blevins’ work on *The Diary of Martha Ballard*, and Robert Nelson’s “Mining the *Dispatch*”. While all effective in their conclusions, each speaks to its own content analysis more than they speak to innovations in pedagogy, approach, and methodology within LDA. Previous events that have focused on topic modeling have been dominated by workshops organized by computer scientists and information retrieval specialists working in natural language processing. These workshops tend to provide low-level technical explorations of particular machine learning approaches, which are obviously not tailored to the training or expertise of general humanities audiences. When the concerns of humanists do intersect with these events, it is often through presentations by computer scientists using humanities’ derived corpora. An example of these phenomena is the 2010 workshop in Natural Language Processing Tools for the Digital Humanities presented at Stanford University during the annual Digital Humanities Conference. Taught by Christopher Manning, a computational scientist, the workshop was a “survey what you can do with digital texts, starting from word counts and working up through deeper forms of analysis including collocations, named entities, parts of speech, constituency and dependency parses, detecting relations, events, and semantic roles, co-inference resolution, and clustering and classification for various purposes, including theme, genre and sentiment analysis. It will provide a high-level not-too-technical presentation of what these tools do and how, and provide concrete information on what kinds of tools are available, how they are used, what options are available, examples of their use, and some idea of their reliability, limitations, and whether they can be customized.” Significantly, this effort to “empower participants in envisioning how these tools might be employed in humanities research” did not close the feedback loop to computational science to imagine how natural language processing tools, including topic modeling software, can be improved to deal with humanities research questions. When humanists are interacting with topic modeling approaches, it is often as an uncritical consumer rather than as an engaged critical applied theorist.

**History and duration of the project:** This workshop has received no previous support. Preliminary research on topic modeling, LDA, and the humanities has been undertaken by Primary Investigator Travis Brown, Research and Development Software Development Lead

at the Maryland Institute for Technology in the Humanities, for a one-year period prior to this application. Brown has been engaged with national dialogues about topic modeling undertaken by computer scientists and information retrieval specialists and has also participated in humanists' discussions of topic modeling via his roles at the Walt Whitman Archive and MITH. He is, as a result, uniquely positioned to facilitate the cross-fertilization process. University of Michigan Graduate Student Sayan Bhattacharyya, and UMD Graduate Student Clay Templeton, who have served as Interns in topic modeling at MITH via an Institute for Museum and Library Services Internship grant in Summer 2011, will aid him. Through their work on Woodchipper, a visualization tool for humanities usage that allows the user to search and select text from participating collections and display relationships among texts, Bhattacharyya and Templeton have aided Brown in identifying thematic areas where cross-fertilization of knowledge about topic modeling needs to occur between humanists, computer scientists, and information retrieval specialists. **Work plan:** The workshop will be organized into three primary areas: 1) an overview of how topic modeling is currently being used in the humanities; 2) an inventory of extensions of the LDA model that have particular relevance for humanities research questions; and 3) a discussion of software implementations, toolkits, and interfaces. Each area will be covered in a two-hour long session with two or three individual speakers giving 30-minute presentations. The initial overview will explore examples of topic modeling approaches currently being used in text analysis projects in the humanities. Potential speakers include, but are not limited to: Matt Jockers, Sharon Block, and Robert Nelson. The overview of extensions will cover a range of variants of the widely-used Latent Dirichlet Analysis topic model that are able to take into account time, geography, and other information about the documents being analyzed and their context, and may include speakers such as Jordan Boyd-Graber, Doug Oard, and Jason Baldridge. The final implementation session will focus on the development and explication of tools such as the Machine Learning for Language Toolkit (MALLET). Potential speakers include, but are not limited to, David Mimno and Taesun Moon. Each area session will culminate in a 30-minute exercise to identify areas of overlapping interest for further development. The workshop will close with an additional 45-minute session that will focus on extrapolating from the individual sessions to a larger understanding of how topic modeling approaches can advance humanities scholarship. **Staff:** The proposed workshop on topic modeling is fortunate to benefit from a variety of substantial relationships at the University of Maryland and MITH. Core project staff will include: Travis Brown, lead Research and Development Software Developer at MITH, who will lead the project and supervise all project activities; University of Michigan Graduate Student Sayan Bhattacharyya and UMD Graduate Student Clay Templeton will work with Mr. Brown to develop an appropriate cyber-environment to gather all associated publications, software, and presentation materials for workshop events; Dr. Jennifer Guiliano, Assistant Director of MITH, will provide logistical support for all workshop related activities including handling all local arrangements and handle fiscal reporting activities; Emma Millon, Community Lead at MITH, will be responsible for all community outreach including distribution of the workshop solicitation, documenting workshop activities via social media, and aiding Brown in completing the white paper. **Final product and dissemination:** We will document publicly the workshop and all associated presentations thereby encouraging other researchers to join our community, benefit from our investment of resources, and extend the discussions related to topic modeling. Using twitter, blogs, and video feeds, we will provide synchronous and asynchronous methods of workshop involvement. By utilizing the workshop website as an opportunity to create a public presence around topic modeling and the humanities, we hope to extend our impact by providing a space for scholars to engage pre- and post-workshop. To aid this, we will release a reflective white paper at the end of the grant documenting the various sub-areas within topic modeling in the digital humanities and attempt to extrapolate potential understanding of how topic modeling efforts can extend itself into humanities scholarship via specific recommendations for further development.

**Citation:**

Bajcsy, Peter, Kevin Franklin, Karen Fresco, Jennifer Guiliano, Anne D. Hedeman, Rob Kooper, Robert Markley, Tenzing Shaw, and Michael Simeone. "Digging into Image Data to Answer Authorship-Related Questions." National Science Foundation, Grant Submission, University of Illinois, Urbana, Illinois, 2009.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Sheffield, UK  
Humanities Research Institute  
Department of French, School of Modern Languages and Linguistics  
University of Illinois at Urbana-Champaign, USA  
National Center for Supercomputing Applications  
Institute for Computing in Humanities, Arts, and Social Science (I-CHASS)  
Department of French  
Department of English  
Art History, School of Art and Design  
Michigan State University, USA  
MATRIX: Center for Humane Arts, Letters, and Social Sciences  
Department of Computer Science and Engineering  
MSU Museum and Department of Art and Art History  
Alliance of American Quilts, North Carolina, USA

**Team members:**

Peter Ainsworth  
Department of French, School of Modern Languages and Linguistics and Humanities  
Research Institute, University of Sheffield, UK  
Simon Appleford  
Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of  
Illinois at Urbana-Champaign, USA  
Peter Bajcsy  
Image Spatial Data Analysis Group, NCSA, University of Illinois at Urbana-Champaign,  
USA  
Steve Cohen  
MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University,  
USA  
Wayne Dyksen  
MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University,  
USA

Kevin Franklin

Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of Illinois at Urbana-Champaign, USA

Karen Fresco

Department of French, University of Illinois at Urbana-Champaign, USA

Matt Geimer

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Jennifer Guliano

Institute for Computing in Humanities, Arts, and Social Science (I-CHASS), University of Illinois at Urbana-Champaign, USA

Anne D. Hedeman

Art History, School of Art and Design, University of Illinois at Urbana-Champaign, USA

Anil K. Jain

Department of Computer Science and Engineering, Michigan State University, USA

Rob Kooper

Image Spatial Data Analysis Group, NCSA, University of Illinois at Urbana-Champaign, USA

Mark Kombluh

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Marsha MacDowell

MSU Museum, Department of Art and Art History, Michigan State University, USA

Robert Markley

Department of English, University of Illinois at Urbana-Champaign, USA

Michael Meredith

Department of French, School of Modern Languages and Linguistics, University of Sheffield, UK

Amy Milne

Alliance of American Quilts, North Carolina, USA

Dean Rehberger

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Justine Richardson

MATRIX: Center for Humane Arts, Letters, and Social Sciences, Michigan State University, USA

Tenzing Shaw

Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, USA

Michael Simeone

Department of English, University of Illinois at Urbana-Champaign, USA

## **Acknowledgments**

The project is supported by the National Science Foundation and National Endowment for the Humanities from the United Kingdom and the Social Sciences and Humanities Research Council from Canada via a Digging into Data Challenge Grant Award. The material presented is based upon work supported by the National Science Foundation under Grant No. 10-39385.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

## Introduction and Project Aims

For more than 40 years, humanities scholars have used computational analysis to help resolve issues of authorship. Through stylistic and linguistic analysis, researchers have puzzled out answers to questions that range from who wrote *The Federalist Papers* and to who collaborated with Shakespeare on *Henry VIII* and *Pericles*. While determining a writer's "genetic fingerprint" is a difficult task, the wealth of scholarship and algorithms that have developed around printed textual analysis promises to help solve a number of vexing authorship issues as well as expand our knowledge of the written arts. However, in the area of visual arts, computational analysis of authorship has not made the same inroads. To do authorship studies of visual works, scholars must often do painstaking point-by-point analysis of small sets of 2D images of the objects. This work becomes all the more difficult when dealing with cultural artifacts such as quilts, maps and medieval manuscripts that often have corporate and anonymous authors working in community groups, guilds, artisan shops, and scriptoria. Beyond the difficulties of authorship attribution, larger important humanities questions about the influence and migration of artistic elements and patterns become all but impossible to assess when large datasets require individual scholarly inspection of each image. To this end, we propose to address authorship and the corresponding image analyses leading to computationally scalable and accurate data-driven discoveries in image repositories.

This effort will utilize three datasets of visual works -- 15<sup>th</sup>-century manuscripts, 17<sup>th</sup> and 18<sup>th</sup>-century maps, and 19<sup>th</sup> and 20<sup>th</sup>-century quilts. Overarching humanities research questions emerge from these groups of works, such as how visual and production styles reflect regional tastes or historical moments, how traumatic historical events manifest in cultural production, and how artifacts reflect and influence relationships between cultural groups. Together these works present a range of complex authorial relationships and a strong base for advancing knowledge on the research problems inherent to scalable, automated image analyses. Open research problems are divided below into artistic, scientific and technological questions based on the specific datasets that elicit those questions. We expect these questions will be useful across the work of all three groups.

For the 15<sup>th</sup>-century manuscripts, Froissart's Chronicles, the *artistic questions* include: Where and by whom were these manuscripts created? How does a manuscript reflect the tastes of the particular region and historical moment to which it belongs? What does the codicological evidence—scribal hands, catchwords, page layouts, artistic styles in the miniatures and marginal decoration—suggest about book production in this period? The *scientific questions* for Froissart's Chronicles ask: Since these manuscripts were made during the Hundred Years' War, what was the impact of war on culture as measured by the various aspects of these manuscripts, e.g., evidence of patronage? How do they reflect contacts between the cultures of France and England? How do they reflect the ideology of chivalry or the concept of history? The questions for these medieval manuscripts are related to: (a) studying the composition and structure (codicology) of the manuscripts as cultural artifacts of the book trade in later medieval Paris; and (b) identifying the characteristic stylistic, orthographic and iconographic 'signatures' of particular scribes and artists and their collaborators who contributed to the illustration and decoration of the volumes, through the use of image recognition and data mining techniques. A further potential output from identifying scribal hands using image analysis techniques is a process that can transcribe the text from the images, a task that is currently done manually by skilled scholars. Thus not only would the content be subjected to analysis but it might also be possible to process it to allow scholars to perform further text-based mining (although not as part of this proposal) on the previously untouchable textual corpus that is locked away as pixels in an image.

The 17<sup>th</sup>- and 18<sup>th</sup>-century maps come from atlases by Joan Blaeu and Herman Moll (original atlases and digital scans held at the University of Illinois Library). The *artistic questions* for these maps include: What characteristics distinguish individual and corporate groups of artists and engravers? Criteria such as color palette, graphic representations of ships, shading of coastlines, and fonts can be considered as distinctive traits that identify both a) particular artists and engravers, b) the corporate styles developed by the Blaeu family in 17<sup>th</sup>-century Amsterdam (Joan was the son of Willem Blaeu, who founded the largest mapmaking engraving and publishing house in the world) and by Moll and his collaborators who adapted Dutch conventions of mapmaking for English audiences in the early 18<sup>th</sup> century, and c) national styles of depicting specific geographic and manmade features (cities, fortifications, trade centers, etc.). The *scientific and technological questions* are: Do specific maps show a more detailed geographical and/or climatological knowledge in representations of coastlines and harbors? Or navigable rivers? Or shoals and sand bars that pose dangers for ships? Or mountain passes that indicate potential routes for exploration and trade? The scientific and technological questions both influence and are influenced by the artistic

questions. In particular, engravers develop specific artistic techniques for representations that were essential for ships' captains, navigators, and merchants who used published maps to sail often unfamiliar and dangerous waters in South America, Asia, and the Pacific (see Appendix D). Maps therefore negotiate among art, science, trade, and politics, and determining the principles that allows researchers to distinguish among different maps and mapmakers will aid scholars working in the history of science and cartography, art, literary studies, colonial history, and economic history.

For 19<sup>th</sup>- and 20<sup>th</sup>-century quilts, *artistic questions* include: What are the distinct characteristics of an individual quiltmaker or relevant quiltmaking group's choices of pattern selection, fabric and color choices, execution of measurement, layout, needlework and craftsmanship of the pattern design, and, most interestingly, original deviations from traditional patterns? Published quilt patterns became much more common starting in the late 1800s, when certain pattern designers mass-produced their patterns and disseminated them through ladies magazines, and later in syndicated newspaper columns. Geographically dispersed quiltmakers who were exposed to this media began gaining new patterns and pattern ideas. Thus, in a large test bed of documented historic quilts, the societal rise and influence of mass media should be seen through the proliferation of quilts that execute patterns disseminated through syndicated columns. The *scientific questions* include: Can the quilts created by quiltmakers from a cloistered family, community, ethnic, or religious group at a particular time period be differentiated from those of other communities, especially those more exposed to mass media? If so, can changes in the community's participation in mass culture be found through changes in quiltmaking styles? Can a resurgence or interest in a particular historic cultural community's quiltmaking styles be found in quiltmaking a century later? To what extent are quilts made by one Amish family in the 19<sup>th</sup> century similar or dissimilar to those made by urban quilters in the same time period? Does this change over time? Or, from an even more fine-grained perspective, do we find more or less divergence in quilts from the North and from the South? To what extent are quilt patterns regional and to what extent national? Does this change over time? A major theme in American cultural history is the eclipse of regional cultural differences during the 20<sup>th</sup>-century. Can we test that hypothesis by looking at quilts? Can we use the Quilt Index dataset to measure the impact of traumatic historical events—say 9/11 or Pearl Harbor—on American culture? Do we see a measurable change in imagery, colors, or composition after such events?

While identifying distinct characteristics of artists is time-consuming, computer-assisted techniques can help humanists discover salient characteristics and increase the reliability of those findings over a large-volume corpus of digitized images. Computer-assisted techniques can provide an initial bridge from the low-level image units, such as color of pixels, to higher-level semantic concepts such as brush strokes, compositions or quilt patterns. The technological questions are related to the design of algorithms that can extract evidence at the low-level image units that could be aggregated into higher-level semantic concepts and support humanists in image understanding and authorship assignment. The further technological questions are about the statistical confidence of authorship hypotheses obtained by processing volumes of images that could not have been visually inspected with the current human resources within a reasonable time frame. How to extract knowledge about authorship and how to increase our confidence in the characteristics of authorship are the key technological questions.

### Problem Description

Based on the artistic, scientific or technological questions, we formulate and address the problem of finding salient characteristics of artists from two-dimensional (2D) images of historical artifacts. Given a set of 2D images of historical artifacts with known authors, we aim to discover what salient characteristics make an artist different from others, and then to enable statistical learning about individual and collective authorship. The objective of this effort is to learn what is unique about the style of each artist, and to provide the results at a much higher level of confidence than previously has been feasible by exploring a large search space in the semantic gap of image understanding.

### Motivation

Currently, humanists must look at images of historical artifacts to determine distinct characteristics of certain individual (or groups of) miniaturists and map engravers, scribes, quilters, and so on. Such visual inspection involves identifying objects in 2D images, recognizing specific types of objects, discriminating differences among those objects, classifying realizations into groups with similarities, building cumulative evidence over multiple groups of objects with similarity in realization, and assigning a authorship based on temporally evolving expertise in visual inspection. For example, to assign a label of an artistic hand to an illustration in Froissart's Chronicles, we would first

identify objects such as boats, castles, crowns, faces, group of knights, horses, landscapes, skies, spears, tents, towns and water. Next, we would look for and identify the discriminating differences in all found instances of these objects and group the objects based on similarities. Finally, using meticulous book-keeping of all groups with similarities, we would build a mapping between the groups of classified objects and the potential assignment of authorship. This manual process is very labor-intensive and cannot be scaled up to large volumes of digital artifacts. In addition, the salient characteristics (a collection of discriminating differences) per artist are described at a high semantic level, which makes it difficult to automate the discovery process. Thus, there is a need to explore the semantic gap in image understanding and to establish the mappings between pixel level image properties and the higher-level abstract image descriptions.

### Data Repositories and Selection Criteria

*Data repositories:* The primary datasets that we propose to use for this research include:

1. Nine complete Froissart manuscripts from the 15<sup>th</sup> century that have been digitized to similar standards and quality (see Appendix C). These are: Toulouse, Bibliothèque d'Etude et du Patrimoine MS 511, Besançon, Bibliothèque d'Etude et de Conservation MS 864 & MS 865, Stonyhurst College MS 1, Brussels, Bibliothèque Royale MS II 88, MS IV 251 tomes 1 & 2, and Paris, BnF MS S français 2663 and 2664. We are currently seeking funding to add two further complete manuscripts to this dataset: Pierpont Morgan Library MS M.804, and British Library MS Royal 15 E.VI. The current collection of 15<sup>th</sup>-century manuscripts consists of over 6,100 images mainly at 500 DPI, hosted on a federated Storage Resource Broker (SRB) facility between UoS and UIUC using a web-front end collaboratively developed by the two sites (see <http://cbers.shef.ac.uk>). The images can also be retrieved from the SRB system via an API that provides direct access to the image dataset within a programming environment such as the Image To Learn toolset (see technical methodology section).
2. Details on the 17<sup>th</sup>- and 18<sup>th</sup>-century map collections: the University of Illinois Library holds a 1664 Blaeu Atlas and over twenty of the Atlases published by Herman Moll in the early 18<sup>th</sup> century, as well as digital scans of the maps for this project (see Appendix D). These atlases include hundreds of additional maps, and the algorithms developed by this project can be applied to the thousands of pre-1800 maps that are gradually being digitized by libraries across the world. There are currently no systematic means of determining authorship for many of these maps, and the open source software developed by this project will help to encourage more digital scans of these rare and valuable but understudied resources.
3. Details on 19<sup>th</sup>- and 20<sup>th</sup>-century quilt images: the Quilt Index (a partnership of Michigan State University and the Alliance for American Quilts) contains images and detailed information on nearly 25,000 quilts, which will grow to 50,000 by the end of the grant period (see appendix E). The quilts, dating from the 1700s to the present day, are mostly American in origin though the Index will expand to include international collections in the future. Access images (550 pixel-wide JPEG files 72-150 ppi resolution) have been contributed by museums, libraries and documentation projects for education and research use. The set is hosted in MATRIX's open source digital repository, KO RA, and available at [www.quiltindex.org](http://www.quiltindex.org). Many thousands of styles and quilt makers are represented in this dataset as well as a range of image quality depending on original photography. For this project we have selected groupings to address four aspects of authorship: Illinois Amish family quilts from the 1800s, 1930-era Detroit quilts of Mary Schafer (who developed a very distinctive border style), typical turn of the century "crazy" quilts by Iowa quilt maker Lottie Enix, and quilts made by multiple quiltmakers using a published 1930s pattern by artist Eline Foland. Determining salient characteristics of colors, shapes, borders, layouts and patterns with these four distinct groupings will be important to automated clustering which will uncover important new similarities and dissimilarities for a broad range of humanities analysis.

*Selection of data repositories:* Given the overarching goal of understanding characteristics of authorship, the proposed framework should consist of generic image analysis algorithms that could be used or adapted for use on other projects and many other datasets. We have selected datasets that represent three different historical periods and three different media but that raise analogous problems in determining authorship. The purpose of choosing such a variety of datasets is to show how seemingly different humanities research questions can share software and resources. The diversity across the three major datasets permits us to consider the computational scalability and broad applicability of the image analysis algorithms; hence we will not be producing methodologies that are only suitable to one specific type of dataset; they will have a much wider impact and use. The different datasets will further

foster integration and evaluation of algorithms so that common parts across many datasets as well as dataset specific parts of algorithms would be well understood. In addition, work developed across three or more cognate projects is certain to reinforce critical mass and to establish a creative dialogue; solutions that may seem relatively obvious to one project team may prove to be a revelation to another team in the consortium.

### Project Methodology (Approach)

We propose to break down the computing problem of discovering salient characteristics into three low-level semantic components characterizing image content: (1) image representations, (2) feature descriptors, and (3) machine learning methods and similarity metrics for assignments of authorship (Appendix B illustrates a diagrammatic outline of this process).

(1) Image Representation: the image representations refer to various ways in which digital images could represent the information about physical objects. The representations include Color spaces (e.g., RGB, HSV, YUV, CIE) [1], Frequency transforms (e.g., Fourier, Hough or digital cosine transform), Special transforms (e.g., Gabor filters, co-occurrence matrices), Decomposition transforms (principal components, wavelets) and Edge transformations (Sobel, Canny, Robertson, etc. [2]). While there have been studies of what representations are close to human perception following Gestalt psychology principles [3], it has not been established how the image representations map towards discriminating artists and to higher-level semantic descriptions. We plan to explore the search space of the above image representations.

(2) Feature descriptors: Once an image representation has been selected, there is a need to describe properties of image pixels (called features) that capture local and global, deterministic and statistical, spatial and spectral image characteristics. The extraction of features can be specifically designed to focus on color, shape, texture, or motion properties. We plan to explore the search space of the most common features including 1D vector of values, color histogram, shape context [4], or texture descriptors (e.g., extracted from co-occurrence matrices) [5].

(3) Machine learning methods and similarity metrics for assignments of authorship: Given a set of features and classification labels of authorship, there exist many machine learning methods that would generate data-driven models (mappings) to convert input features into a desired set of labels. The data-driven models compare input features using similarity metrics and try to find parameters of a model that would predict the authorship labels accurately. The positive yield of models is evaluated based on the number of correctly classified instances of input features. We plan to explore the search space of multiple machine learning methods including K-nearest neighbors, support vector machine (SVM), artificial neural network (ANN), decision tree, and K-means [6].

Our approach is to explore a large dimensional space consisting of all possible combinations of image representations, feature descriptors, supervised machine learning methods and their parameters in order to select the most salient characteristics per artist. These characteristics per artist are selected based on the accuracy reported by supervised machine learning methods that compare predicted authorship assignment using the data-driven models with the provided authorship labels. The result of such extensive searches would lead to an n-tuple that provides the highest discrimination with reference to two artists. For instance, let us assume that the n-tuple found consists of (a) hue color channel in HSV image representation, (b) frequency of occurrence of each hue value – hue histogram, and (c) similarity of hue histograms measured by chi-squared error and aggregated into groups with similar features using three nearest neighbors. Then, a humanist could interpret the discriminating characteristics of two artists to be a hue component of image colors, a statistical distribution of hue variations in image pixels, and neighboring similarity of hue distributions in the space of all possible statistical distributions. Thus, visual inspections by a humanist would be assisted by a computer-driven recommendation to focus on a hue component of color images and the similarity of hue distributions in images (or the similarity of hue value frequencies across images). This would reduce the search time of a humanist and could change the role of visual inspection from searching to verification and validation. Furthermore, the images would be delivered for visual inspection in the appropriate representation (e.g., hue channel and its hue histogram) rather than leaving a humanist to recover the hue representation from another color space representation by color transforming images inside of his/her brain.

Similarly, these pair-wise (artist-to-artist) analyses would lead to a matrix of discriminating characteristics that could be summarized and presented to a humanist. The summaries provide computer-assisted input into research questions about what salient characteristics of an artist dominate within a group of artists, a school of artists

or a community of artists. Furthermore, they would be tremendously useful in forensic studies when unseen images are presented to determine authorship.

Essentially, and in respect to determining scribal hands, this scientific methodology would help scholars to identifying recurring tell-tale signs pinpointing the work of Scribe A, B or C by providing answers to our scholarly questions such as which forms, or combinations of letters (e.g., ligatures such as 'ct' or 'br'), do our algorithms reveal as being key to distinguishing between A, B and C?

### Computational Requirements

The proposed approach is clearly computationally intensive due to the huge size of the search space for selecting the optimal triplet (the most salient characteristics represented by image representation, features and machine learning methods). The dimensionality can be estimated by multiplying the number of image pairs to be evaluated times the number of image representations times the number of features times the number of machine learning metrics times the number of cross validations. For example, computing the optimal triplet that discriminates between two artists, each represented by 10 images, over 2 color spaces (HSV, H, S, V, HS, HV, SV, RGB, R,G, B, RG, RB, GB), 1 feature type (histogram of each band with the number of bins varying from 100 to 255), one machine learning method (k-nearest neighbors with the k parameter taking values 1, 2,3,4,5) and 5-fold cross validation requires  $(10+1)*5$  of image pairs to be evaluated times 14 image representations times 155 features times 5 machine learning variations times 5 cross validation evaluations. This number is equal to 2,983,750 computations with many floating point operations during accuracy evaluations of machine learning models. Clearly one feature type and one machine learning method will not be sufficient to capture the spectrum of underlying image characteristics that discriminate artists. We anticipate computations to number about 300 million after adding variables on each dimension and reducing the number of parameters based on initial optimizations (e.g., the number of histogram bins).

### Development Methodology & Standards

The Image to Learn (IM2 Learn) toolset<sup>1</sup>, developed at NCSA, provides a suite of image transformation functions that will be applied to our dataset to support the first phase of our investigations. IM2Learn additionally provides an API to facilitate the development of other image transformation algorithms that do not currently exist within the package. This standard API will be used by the project partners while developing further functions to ensure cross-compatibility across the sites.

The feature description catalogue (phase 2) will be recorded using an XML database and document type definition (DTD) that will be developed in consultation between the technical developers and scholars to meet the needs of the project from both perspectives. This will provide a structured document that is both human and machine readable that will be directly input into phase 3.

The machine learning component of our methodology (phase 3) will make use of the open source Weka software<sup>2</sup>, developed by the University of Waikato, which provides a collection of machine learning algorithms and a platform to develop additional functionality to suit this proposal. However, unlike the image transformation phase and the use of IM2Learn, our project may not be best served by a direct uptake of the Weka interface due to an increased complexity in operations, particularly since one of the aims of this initiative is to scale-up the algorithms to work on high-performance computing grids. The technical project team will therefore meet to decide a suitable framework, also taking into account previous decisions from the earlier stages to ensure the process is both streamlined and coherent while the partner sites collaborate on aspects of functionality.

The algorithms contained within the IM2Learn and Weka packages will be enhanced by the inclusion of new and existing code from the project partners converted into the common framework/interface. This activity will be supported by bringing together the algorithms that each partner has developed in past projects which relate to this application. Furthermore, to facilitate scalability, we anticipate that the algorithms that feature in our final pipeline will need to be written and optimized to allow large-scale parallelism. The new algorithms developed within the scope of this proposal will be made available as open source in line with the exit strategy outlined below.

---

<sup>1</sup> <http://isda.ncsa.uiuc.edu/Im2Learn>

<sup>2</sup> <http://www.cs.waikato.ac.nz/ml/weka>

Documentation on the framework and an evaluation of the interfaces, protocols and standards used during the development stages will be published as part of the final report. This will include particular references to the team success in achieving a platform for working collaboratively on interoperable code across geographically remote sites.

### Environmental Scan

Based on our knowledge, there has not been an effort in the past such as the proposed one. There are publications referring to forensic studies of individual photographs of arts investigating the authenticity [7, 8], or the use of optics during art creation [9, 10]. Several researchers have explored certain sub-sets of features, for instance, brush strokes [11, 12], for classification purposes. The closest effort to the proposed one is the work of Shen [13] where 1080 classical paintings from 25 artists were analyzed using a collection of color, texture and shape features. However, the search space in [13] is constrained to only searching over one data representation (CIE) and one machine learning method (neural network). In addition, our proposed effort is different not only in the scale of data sets and required scalability of search computations but also in posing a fundamentally different question of finding salient characteristics discriminating two authors and groups of authors at the low image level by computers and then at the higher semantic level by humanists. This approach has been initially investigated by two coPIs of this project at UIUC (Hedeman and Bajcsy) with funding from the NCSA Fellowship program [14].

### Final Product and Deliverables

The final product will primarily consist of (a) data about salient characteristics of an artist with respect to another artist and with respect to a group of artists, and (b) software for obtaining salient characteristics. The data could be viewed as evidence supporting a authorship assignment based on (1) image-derived primitives including image representation, image features and machine learning model for assigning authorship, and (2) human-defined semantic descriptors of unique characteristics that map into a combination of multiple image-derived primitives.

The final product would be used immediately by scholars for book trade studies, for understanding engravers and cartographic artists, and for addressing questions about collective and individual authorship of quilts. Additional likely users of the final product would be the researchers and students at UIUC, MSU, University of Sheffield, and other universities for educational purposes. We foresee other humanists wanting to know about the authorship of all kinds of collections and using the software framework and the algorithms. The algorithms and software developed will have appropriate technical and user documentation and all content made publicly available as open source for non-commercial use. The final products will demonstrate to a wide audience how large-scale data repositories of cultural heritage materials can change the nature of research for the humanities and social sciences.

### Dissemination

Websites will be created at each of the partner sites at the start of the project to introduce and publicize the work. These sites will act as primary points of contact for project information, offering up-to-date examples of results and blogs (which will be used to inform the final report). The websites will host the final deliverables and include the source code and documentation developed during this initiative.

A workshop in Sheffield is planned to take place around month 9 to bring together other expert scholars in the field of authorship from around the UK. Its primary purpose is to engage with the wider community and understand more broadly how other experts interpret authorship and how the preliminary project results might help their work. We also anticipate that the workshop will influence our choices of image representations during the machine learning phase. The engagement we accomplish during this workshop will also disseminate the work by demonstrating its potential to a wider audience, who themselves have a specific interest in authorship.

Further avenues for dissemination and publication that the UK team will target include the All Hands and Digital Resources for the Humanities and Arts conferences which showcase e-Science tools and methodologies and the Digital Humanities Quarterly journal. Other workshops and conferences will also be attended to disseminate and publish this work. Members of the UIUC team will present status reports on the project at national and international conferences in art history, cultural history, and medieval and eighteenth-century studies.

### Project History

These three partners each have considerable experience with their own datasets represented, and have established working relationships on imaging issues prior to the development of this proposal. In April 2009 members of each partner site were part of an NSF workshop, "Imaging and Image Analyses Applied to Historical Objects," at

the University of Illinois which addressed the process of going from physical historical objects to digital historical objects available via the Internet for educational and research purposes. The overarching theme of the workshop was to understand the challenges associated with imaging and image analysis that are inherent in this process.

The Sheffield project director (PD) is close to completion of the AHRC-funded "Online Froissart" project. The project will deliver an online facsimile/transcription and translation from Books I-III of Froissart's Chronicles, based on digital surrogates of original manuscripts and using XML. The "Online Froissart" will therefore help reinforce mark-up techniques and standards for the present proposal as well as provide knowledge about scribes and artistic hands; research tools emerging from the project, including image-processing algorithms, will also feed into the Digging into Data proposal. The Sheffield PD has also led on the EP SRC-funded "Pegasus" project, which considered grid-enabled interfaces [15] for sharing and displaying in real-time online image datasets and virtual exhibitions to a distributed audience; this was done in partnership with Urbana and used the "Virtual Vellum" software from Sheffield.

With the support of a Faculty Fellowship (2008-09) from NCSA, Co-PIs from the University of Illinois (Hedeman and Bajcsy) began work on developing cyber tools for analyzing the visual imagery embedded in the corpus of Froissart manuscripts (a corpus that will ultimately include the Shrewsbury Book and Morgan MS), in order to provide insights into both the artists' contributions to the construction of these specific books, and more broadly, the functioning of the medieval book trade. The Sheffield and Urbana team have also worked closely on other grant applications including a pending JISC/NEH digitization proposal. NCSA researchers have also explored preliminary automated pattern analysis of "crazy" quilts, which will be part of the dataset and questions addressed in this project. Finally, NCSA and ICHASS have been actively supporting the 18<sup>th</sup> Connect project with computational resources and expertise needed for pre-1800 optical character recognition.

#### Time Plan and Project Management:

- Month 1: Data and initial source code sharing platform established between all partners with all relevant material uploaded. Project meetings will be held over Access Grid. Project websites launched at all three sites. Consortium agreement (JISC requirement) including any new responsibilities, data policies, finalized evaluation plan, etc. arising from first round of project meetings.
- Month 2: Scholars and developers discuss base-line algorithms and selection of training data
- Months 3-4: Integration of existing algorithms and alpha development of new image transformation code from initial conversations. The alpha versions of new code will be rapidly developed (i.e. not optimized) as a means of getting results quickly to determine how useful the algorithm will be long-term.
- Month 5: Scholars determine appropriate feature sets based on transformations with discussions with developers and DTD for XML database decided.
- Month 6: Optimization, modification and validation of image transformation code for large-scale processing.
- Months 7-8: Apply machine learning algorithms to test dataset to determine suitable discriminative n-tuples suitable for authorship classification. Project exposure to classroom students at UIUC, MSU and UoS
- Month 9: Sheffield workshop to discuss results and authorship issues among larger scholarly community. Analyze results to prune misleading representations. Optimize algorithms for large-scale processing.
- Month 10: Re-apply machine learning algorithms to test dataset to determine validity of changes.
- Months 11-12: Final data preparation, computational resource configuration and data processing of complete dataset collection. Presentation of the work at the eScience meeting.
- Months 13-14: Documentation of algorithms and technical processes. Scholarly interpretation of results. Draft final report. Incorporating the project results into teaching materials.
- Month 15: Completion of documentation, final reports (including financial statement) and website updates. Dissemination of project results to classrooms outside of the three partnering universities.

*Note: During development of algorithms, the developers and scholars will continue to work closely together; the milestone months indicated above will be used to consolidate progress; formal meetings will be held to monitor progress across all partners.*

The three project directors from each partner site (Ainsworth, Bajcsy & Rehberger) and project managers (Guiliano & Richardson) will meet formally on a quarterly basis to monitor and report overall progress and respond to unexpected issues that risk causing deviation from the time plan. These meetings will be conducted over access grid/teleconferencing. Each site project director has responsibility for overall management of his/her research teams

(as per funding bodies) and reporting back to the funding councils. The research teams primarily include scholars and technical staff – the project directors also fall into one of these categories – and are responsible for undertaking the work as outlined above.

The project team has been selected across the partner sites to bring together the best cross-disciplinary levels of expertise calculated to deliver the collective goals (see Résumés section for team and individual areas of expertise). Expertise in the different areas will be shared throughout the project to meet the objectives; all technical staff across the sites will work closely together to develop the algorithms, as opposed to each site focusing on a specific dataset. This has the additional benefit of ensuring that work is not duplicated, while cross-site compatibility is assured by the common framework and interfaces that will be used: the Image2Learn API will enforce the standards used to develop the image transformation algorithms; however the machine learning algorithms and optimization steps are more technically involved. Thus to ensure consistency between sites, the Sheffield developer will spend a few weeks at NCSA working with their developers at this stage of the project (approximately month 6/7). During this time, any APIs necessary and methodologies required to run the code across the different sites on different grid infrastructures will be established along with optimization strategies. The coordination of work across the technical staff will be the responsibility of the head technical developer and NCSA project director, Peter Bajcsy. All technical staff will liaise informally via electronic communications.

Dissemination activities will be undertaken by either the research team or specialized personnel depending on how the research teams at each site are organized (see list of participants). Dissemination personnel will be managed by the site PD and overall strategic dissemination policies will be realized across the whole team. Throughout the project all team members will communicate with each other and outcomes will be disseminated throughout the team via the blogs. Formal meeting minutes will be made available on the websites.

As one part of project management, the teams have also discussed budget allocations devoted to the three key components of the project, such as computer science (CS), humanities (H) and dissemination (D). The teams have aligned their resource allocations with the project requirements in order to contribute with complementary expertise. The approximate allocations are (a) JISC funding 80% (CS): 10% (H): 10% (D); NSF funding 73% (CS): 20% (H): 7% (D), and NEH funding 40.1% (CS): 44.6% (H): 15% (D).

### Exit/Sustainability Plan

The project deliverables will be subject to the licensing policies outlined by the Creative Commons initiative, and Illinois open source license<sup>3</sup>, and made publicly available for non-commercial use as open-source, thus fostering adoption within the wider community and use for different datasets. The deliverables will be deposited with JorumOpen (in line with JISC's funding policy) and our project websites for download.

We believe that the scholarly results obtained from the data analysis performed will reach beyond this application with only preliminary findings published in the project lifespan due to the development program and time constraints. The methodologies developed will therefore continue to be used by the team members as new discoveries are found and validated.

From a development perspective, at this stage we do not expect the computer to determine authorship indisputably, but rather to cluster similarities together to allow the scholar to focus on these without trawling through large datasets. The work carried out within this proposal will therefore be used as a platform to bid for further funds from agencies, to allow us and others to continue to explore research directions that this initiative uncovers.

### Risk Management and Intellectual Property Rights

We do not consider there to be any significant risks associated with the proposal. Key staff members are completely committed to the project and we have full rights and permission to use the image datasets in line with the purposes outlined. All existing software we plan to use is open-source. In the event of unexpected changes in staffing, we anticipate that our collective research team and further recruitment will be able to compensate due to the complementary skills that each team member brings to the project.

In terms of the technical validity of this application, a precursor to this work has already been undertaken at NCSA [14], demonstrating promising findings, and thus proof-of-concept, which support the research methodologies set forth in this application. This application will, however, broaden and extend beyond the previous work.

---

<sup>3</sup> <http://www.otm.illinois.edu/node/396>

**Citation:**

Guiliano, Jennifer, Song Wang, and Jun Zhou. "The Sapheos Project: Transparency in Multi-image Collation, Analysis, and Representation." National Endowment for the Humanities, Grant Submission, University of South Carolina, Columbia, SC, 2010.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of South Carolina  
Center for Digital Humanities

**Team members:**

University of South Carolina  
Jennifer Guiliano  
Jun Zhou  
Song Wang  
David Lee Miller  
Aidan Zanders

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

**Citation:**

Guiliano, Jennifer, Song Wang, and Jun Zhou. "The Sapheos Project: Transparency in Multi-image Collation, Analysis, and Representation." National Endowment for the Humanities, Grant Submission, University of South Carolina, Columbia, SC, 2010.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of South Carolina  
Center for Digital Humanities

**Team members:**

University of South Carolina  
Jennifer Guiliano  
Jun Zhou  
Song Wang  
David Lee Miller  
Aidan Zanders

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.



**Center for Digital Humanities**  
University of South Carolina  
Department of English  
Humanities Office Building  
Columbia, South Carolina 29208

David Lee Miller, Director  
dmill1951@gmail.com

September 14, 2010

Re: Digital Humanities Start Up Grant (Level II): The Sapeos Project:  
Transparency in Multi-image Collation, Analysis, and Representation

On behalf of the University of South Carolina, I am writing today to request a change in Primary Investigator for the Digital Humanities Start Up Grant (Level II): The Sapeos Project: Transparency in Multi-image Collation, Analysis, and Representation. Awarded in August 2009, the Sapeos Project began its tenure of the direction of our Associate Director of the Center for Digital Humanities, Randall Cream. Dr. Cream worked in partnership with myself and Dr. Song Wang of the Department of Computer Science and Engineering while supervising the work of three graduate students on the construction of the Saphaeos image collation tool.

In May of this year, Dr. Cream left the University of South Carolina and I began to supervise the project work in conjunction with Dr. Wang. Dr. Jennifer Guiliano, the new Associate Director of the Center for Digital Humanities and Research Assistant Professor in History took over all roles and responsibilities assigned to Dr. Cream as part of this grant and will complete all tasks as projected including finalizing project work by December 30<sup>th</sup> of this year and facilitating a final white paper.

Given the commitment to continuing the work of the project, I write today to request changing the project PI from Dr. Cream to Dr. Guiliano. I do not anticipate any delays or difficulties with the change in Primary Investigators. Attached to this request is Dr. Guiliano's curriculum vitae. Please do not hesitate to let me know if you have any questions.

Sincerely,

A handwritten signature in black ink, appearing to read "David Lee Miller".

David Lee Miller  
Director, Center for Digital Humanities  
University of South Carolina

A handwritten signature in black ink, appearing to read "Jennifer Guiliano".

Jennifer Guiliano  
Associate Director, CDH  
University of South Carolina

# Annual Performance Report

Grant # HD5088009

The Sapheos Project: Transparency in Multi-image Collation,  
Analysis, and Representation

Project Director: Jennifer Guiliano  
University of South Carolina

September 24, 2010

## Annual Performance Report

The Sapheos Project was funded via a Level II: Digital Humanities Start Up Grant to pioneer an innovative product: a digital collation software, prototyped in MATLAB and delivered as an open-source project using C code, that provides both a back-end collation tool and a powerful front-end interface for interacting with large datasets of books. Additionally, our software automatically sections and generates (x,y) coordinate pairs for page images. Written in Java and being ported to MATLAB for C compilation, this software takes images of pages with existing XML markup and inserts (x,y) coordinate pairs into hierarchical elements—lines, paragraphs, stanzas (line groups), and figures—to allow XSL transformations to closely associate textual and image data for users.

Of particular note in this performance review is the change in Project Director necessitated by the departure of Project Director Dr. Randall Cream from the University of South Carolina. Between September 1, 2009 and May 15, 2010, Dr. Cream supervised all project related work through a two-semester sequential course release. This included the hiring of Jarrell Waggoner and Jun Zhou, graduate students in Computer Science and members of the Computer Vision Team, as well as aiding in the creation of the software design plan and implementation of the development of the collation software. On May 15, 2010, Dr. Cream left the University of South Carolina and Dr. David Lee Miller, Director of the Center for Digital Humanities, began to supervise the project work in conjunction with Dr. Wang. Effective July 15, 2010, Dr. Jennifer Guiliano, the new Associate Director of the Center for Digital Humanities and Research Assistant Professor in History, took over all roles and responsibilities assigned to Dr. Cream and held temporarily by Drs. Miller and Wang as part of this grant.<sup>1</sup> Significantly, there were no project delays or changes in the quality of work being completed by grant staff as every effort was made to ensure the continuity of work. Dr. Guiliano is now charged with completing all tasks as projected including finalizing project work, facilitating a final white paper and completing all performance reports.

As outlined below as of August 31, 2010, the Sapheos Project has met all goals within the allotted time and are projected to meet the December 31, 2010 closing date of the grant. These goals with the outlined accomplishments are listed below in table format for ease of reading.

Initial Goals	Actual Accomplishments
Hiring of Graduate Student Developer for Image Processing and Algorithmic Development	Completed September 15, 2009.
Software design and architecture schematic	Initial software design completed September 30, 2009. Architecture schematics projected for completion December 15, 2010.
Implementation of development	Implementation of development began

---

<sup>1</sup> This change in leadership was approved by Michael Hall of NEH on September 17, 2010. Correspondence related to this is available on eGems.

	September 15, 2009 with pre-processing of images. On October 20, 2009, the project continued development along two parallel tracks: automatic overlaying and segmentation/separation of contents.
Separated Printed Contents and Handwritten contents	From November 2009 to March 2010, Jun Zhou researched and tested open source Software for separating printed and handwritten data. By March 30, 2010, Zhou had chosen a package (HandAnalyser) and tested it against an initial pilot set of data.
Extending finding x-y coordinates of text lines and words	Between March 2010 and August 2010, Jun Zhou experimented with using MATLAB to segment lines within image-based texts. On August 29, 2010, Zhou was able to successfully segment the text lines on the initial pilot dataset and outline best practices for this effort.
Image Overlaying	Project staff experimented with both manual and automatic image overlaying. Manual image overlaying was developed prior to the grant award period. Automatic image overlaying began in early August 2010 and is projected for completion at the close of September 2010.
Difference Detection	Initial prototypes for difference detection were completed August 31, 2010 and testing of the prototypes with data and software refinement began September 1, 2010 by Jarrell Waggoner. A completed prototype will be available for download and use on December 31, 2010.
Data visualization	GUI-based data visualization efforts are currently under consideration. Basic visualization of image data was completed in January 2010 and is primarily expressed through data rendering and image processing.
Software testing and Revision	Software testing and revision is underway. It is anticipated that both Waggoner and Zhou will have completely functional prototypes at the close of this grant.
Drafting of Technical White Paper	A draft of the technical white paper is currently being completed with input from Jun

	Zhou and Jarrell Waggoner in consultation with Dr. Wang and Dr. Guiliano.
Hiring of Undergraduate Students for user interface development and testing	Completed August 15, 2010. Shawn Maybay, Bernardo Quibiana, and Maliek McKnight, undergraduate students in Media Arts and Computer Science respectively, were hired in consultation with Jun Zhou of the Computer Vision Team and Center for Digital Humanities.
User Interface Development and Testing	Maybay, Quibiana, and McKnight have begun user interface development and testing and are projected to be completed December 15, 2010. Jun Zhou will coordinate these efforts in conjunction with Dr. Guiliano

Projecting towards our completion of December 31, 2010, we anticipate two software prototypes, a technical white paper, and all performance reporting being completed with no delays or problems. It is our intent to outline in that final report, the best practices for image collation and to underscore the challenges associated with image work: namely, that image content often challenges the ubiquitous nature of software and forces extreme levels of customization and that costs associated with MATLAB and other image processing softwares can be cost prohibitive to humanists.

Final Performance Report

Grant # HD5088009

The Sapheos Project: Transparency in Multi-image Collation,  
Analysis, and Representation

Project Director: Jennifer Guiliano  
University of South Carolina

March 29, 2011

## Final Performance Report

The Sapheos Project was funded via a Level II: Digital Humanities Start Up Grant to pioneer an innovative product: a digital collation software, prototyped in MATLAB and delivered as an open-source project using C code, that provides both a back-end collation tool and a powerful front-end interface for interacting with large datasets of books. Additionally, our software automatically sections and generates (x,y) coordinate pairs for page images. Written in Java and being ported to MATLAB for C compilation, this software takes images of pages with existing XML markup and inserts (x,y) coordinate pairs into hierarchical elements—lines, paragraphs, stanzas (line groups), and figures—to allow XSL transformations to closely associate textual and image data for users.

Between September 1, 2009 and May 15, 2010, Dr. Randall Cream supervised all project related work through a two-semester sequential course release. This included the hiring of Jarrell Waggoner and Jun Zhou, graduate students in Computer Science and members of the Computer Vision Team, as well as aiding in the creation of the software design plan and implementation of the development of the collation software. On May 15, 2010, Dr. Cream left the University of South Carolina and Dr. David Lee Miller, Director of the Center for Digital Humanities, began to supervise the project work in conjunction with Dr. Wang. Effective July 15, 2010, Dr. Jennifer Guiliano, the new Associate Director of the Center for Digital Humanities and Research Assistant Professor in History, took over all roles and responsibilities assigned to Dr. Cream and held temporarily by Drs. Miller and Wang as part of this grant.<sup>1</sup> Significantly, there were no project delays or changes in the quality of work being completed by grant staff as every effort was made to ensure the continuity of work. Dr. Guiliano was charged with completing all tasks as projected including finalizing project work, facilitating a final white paper and completing this final performance report.

The Sapheos Project has met all goals as defined in the initial proposal. These goals with the outlined accomplishments are listed below in table format for ease of reading.

Initial Goals	Actual Accomplishments
Hiring of Graduate Student Developer for Image Processing and Algorithmic Development	Completed September 15, 2009.
Software design schematic	Initial software design completed September 30, 2009.
Architecture schematic	Architecture schematics completed December 15, 2010.
Implementation of development	Implementation of development began September 15, 2009 with pre-processing of images. On October 20, 2009, the project continued development along two parallel tracks: automatic overlaying and segmentation/separation of contents.
Separated Printed Contents and Handwritten contents	From November 2009 to March 2010, Jun Zhou researched and tested open source Software for separating printed and

---

<sup>1</sup> This change in leadership was approved by Michael Hall of NEH on September 17, 2010. Correspondence related to this is available on eGems.

	handwritten data. By March 30, 2010, Zhou had chosen a package (HandAnalyser) and tested it against an initial pilot set of data.
Extending finding x-y coordinates of text lines and words	Between March 2010 and August 2010, Jun Zhou experimented with using MATLAB to segment lines within image-based texts. On August 29, 2010, Zhou was able to successfully segment the text lines on the initial pilot dataset and outline best practices for this effort.
Image Overlaying	Project staff experimented with both manual and automatic image overlaying. Manual image overlaying was developed prior to the grant award period. Automatic image overlaying began in early August 2010 with continuing work in September 2010.
Difference Detection	Initial prototypes for difference detection were completed August 31, 2010 and testing of the prototypes with data and software refinement began September 1, 2010 by Jarrell Waggoner. In November 2010, Waggoner was joined by additional CDH staff and Dhaval Salvi in testing and finalizing the completed prototype. As part of that finalization, Salvi spearheaded the creation of a GUI to integrate the two portions of the prototype into a singular downloadable package.
Data visualization	GUI-based data visualization efforts are currently under consideration. Basic visualization of image data was completed in January 2010 and is primarily expressed through data rendering and image processing.
Software testing and Revision	Initial software testing and revision has been completed with the establishment of a downloadable prototype. The prototype will be available beginning March 31, 2011 at <a href="http://cdh.sc.edu">http://cdh.sc.edu</a> under the Resources menu.
Drafting of Technical White Paper	A draft of the technical white paper with input from Jun Zhou and Jarrell Waggoner in consultation with Dr. Wang and Dr. Guiliano was completed in late November 2010. Drs. Guiliano and Wang offered critical feedback in early December with Jun Zhou taking leadership in revising the technical paper in December. Following the completion of the grant, Zhou working with Salvi augmented the drafted paper with finalized screenshots and code associated with the completed prototype.
Hiring of Undergraduate Students for user interface development and testing	Completed August 15, 2010. Shawn Maybay, Bernardo Quibiana, and Maliek McKnight,

	undergraduate students in Media Arts and Computer Science respectively, were hired in consultation with Jun Zhou of the Computer Vision Team and Center for Digital Humanities.
User Interface Development and Testing	Maybay, Quibiana, and McKnight began user interface development and testing in late fall with final assessments offered to Zhou in late November/ early December 2010. Zhou and Salvi used these assessments in their development of the GUI as well as the final revisions of the software undertaken in December 2010.

Sapheos Project software links the words on a page to the images of the words, figure tags on a page to images of the figures, and transcriptions of marginalia within the XML to images of the handwriting. Our software also automatically sections and generates (x,y) coordinate pairs for page images. Written in Java and being ported to MATLAB for C compilation, this software takes images of pages with existing XML markup and inserts (x,y) coordinate pairs into hierarchical elements—lines, paragraphs, stanzas (line groups), and figures—to allow XSL transformations to closely associate textual and image data that may vary in terms of quality and size for users.

Our most significant issue with regards to our technical goals was linking images to text via XML. In our initial technical plan, we suggested that we would link page images to text encoded in XML to the level of individual characters. This process involved two steps: document image segmentation and injecting the segmentation information (xy-coordinates values) in XML. Here XML file serves not only as a container for the result, but also as a reference to segmentation procedure. For example, `<hi rend="face(ornamental)">` in XML represents a large ornamental letter image that may cross multiple lines. `<lg>` and `<l>` indicate the content is a poem that should be treated as a line instead of a paragraph. We had originally intended to link every word, line, paragraph, and picture in the transcript to its coordinate position in the image to enable other functions like the highlighting on an image of searched-for text. While we were able to roughly segment out words, lines and stanzas, the unusual fonts, warping and degradation present in the document images created a relatively large percentage of error when utilizing fully-automated segmentation to the character level. As a result, we utilized automated segmentation first then undertook a manual correction process. We anticipated a six-week automated/manual linking process, yet even with the automated tool, linking text to image to the level of individual characters proved unfeasible as it took significant labor to complete every single word and line of text. For example, see figure 1 for page image and following segmentation result in xml format.

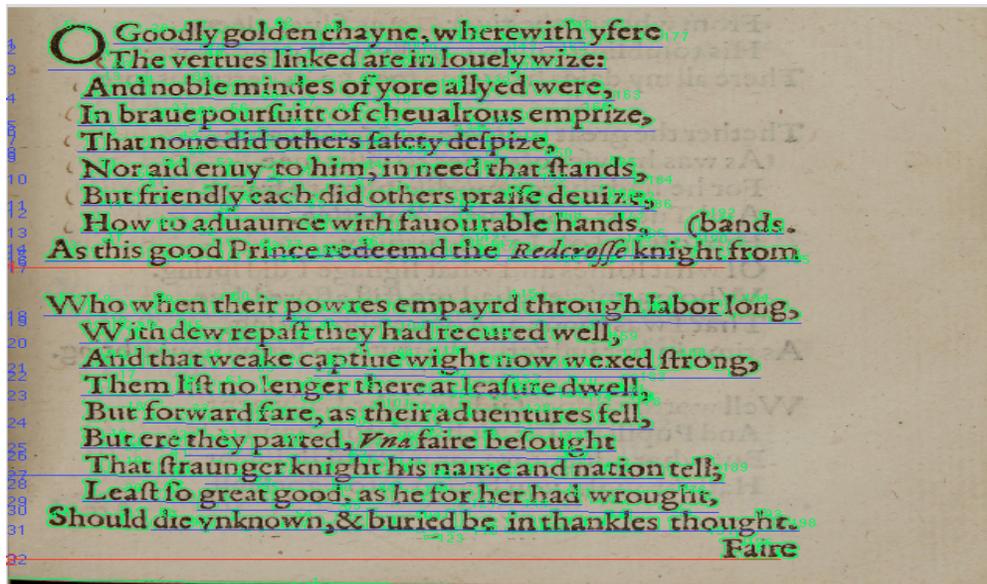


Figure 1: page with segmentation result illustrated in the image.

```

<stanzas>
:
<stanza>
  <stanza number>2</stanza number>
  <x_start>1279</x_start>
  <y_start>1518</y_start>
  <x_end>1495</x_end>
  <y_end>1518</y_end>
</stanza>
</stanzas>
<lines>
:
<line>
  <line number>32</line number>
  <x_start>1279</x_start>
  <y_start>1518</y_start>
  <x_end>1495</x_end>
  <y_end>1518</y_end>
</line>
</lines>
<words>
:
<word>
  <word>202</word>
  <x_start>1482</x_start>
  <y_start>1395</y_start>
  <x_end>1530</x_end>
  <y_end>1440</y_end>
</word>
</words>

```

This page has 2 stanzas, 19 lines and 136 words. The tool gives us 2 stanzas, 32 lines and 202 words. The manual correction for this single page needs to compare every element in the result with the image. After correction, injecting the resultant xml into the existing TEI XML transcription file requires the interjection of every TEI XML tag relating to the result xml element. This involves another round of manual correction if any mismatches exist. For example, if first manual correction segments “&” as a single word, but the transcription put “&” and “buried” as one word then that misidentification needs to be corrected manually. As a result, we believe designing an intelligent automatic linking system is a much larger project than feasible under the auspices of this grant.

Image collation is a process that identifies differences by comparing multiple images from the similar documents. The differences range from semantic difference to character printing error. The existing mechanisms in place for performing collation have traditionally been limited by requiring pristine scanned duplicates of the original witnesses to be compared. Three issues arise with these mechanisms: 1) they are usually obtained by flattening pages of a bound book, risking damage to the binding and adjacent pages; 2) without efficient difference detection system, scholars sometimes need rigorous character-by-character manual inspections, e.g. typographic analysis; 3) or they require complex optical systems utilizing expensive hardware solutions. We have established an automatic image collation or automatic image registration that addresses these programmatic issues. To automate the collation process, precise point matching turns out to be the most critical step in the whole project. For each pair of images, we initially generate thousands of matching points using SIFT algorithm – a state-of-art image processing algorithm to find the matching points. And then we applied a four-tier system of filters to get rid of unnecessary points.

1. Nearest neighbor match filter: good matches should happen within designated neighboring area, e.g., the point at the bottom of template page should never match the point at the top of target page.
2. Duplicate match filter: good matches should be one-to-one matches; this is required by Thin-plate-spline transformation process.
3. Thin plate spline regression filter
4. Local consistency filter

Thin plate spline regression filter and local consistency filter are advanced image processing filters that remove matches affected by noise like document degradation, ink bleeding, etc. The prototype was implemented with MATLAB. We evaluated this prototype by selecting about 172 large size images (1279X2444) and did ten rounds of two-page collation tests. Each test round took about 12 hours running on an eight-core processor MacPro server. All of our tests showed that all differences could be identified correctly. However, false positives were a significant challenge and are addressed below. Constructing a user-friendly interface that can efficiently analyze historical document images was precisely our intention, and we have succeeded in completing this endeavor. Our user interface was built as a standalone JAVA application. As a cross-platform language, JAVA turns out to be the first choice for other popular image systems, i.e. ImageJ, Virtual Light Box. As promised, our application allows a user to load up to four pages. Using the choice buttons at the panel bottom, the user can select and deselect any two of them to collate [figure 2]. The pages can be made translucent in order to give the user a better visual result (differences shown as red marks) [figure 3]. The most important of all, while doing collating, the user does not need to align the images by tedious manual dragging.

All of these are done automatically, even with warping presented in the images. The results are displayed as two pages perfectly overlaid with difference shown as blurs.

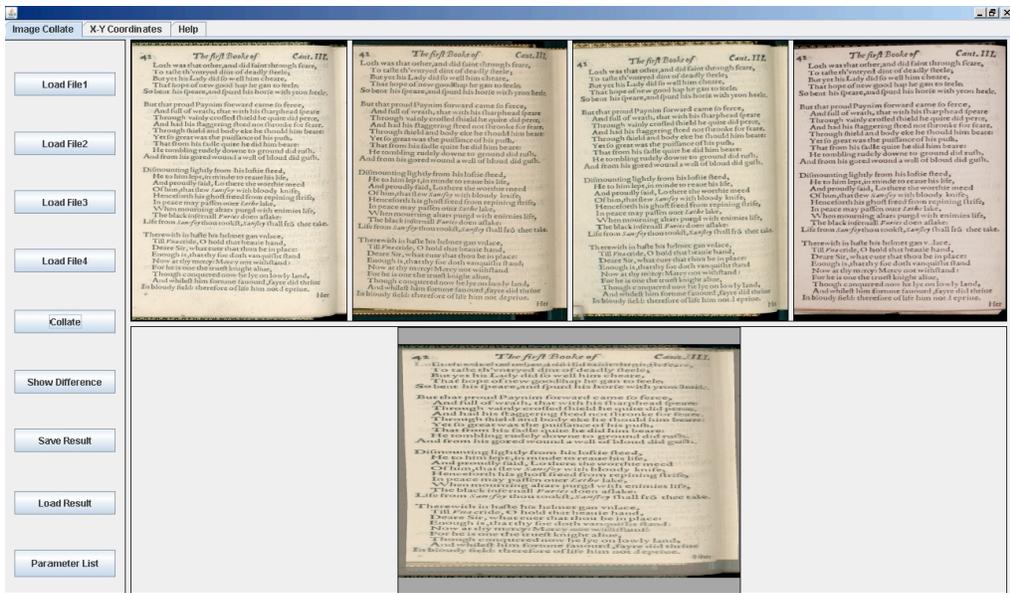


Figure 2: Collating multiple pages, result of two-pages images shown in the bottom right panel.

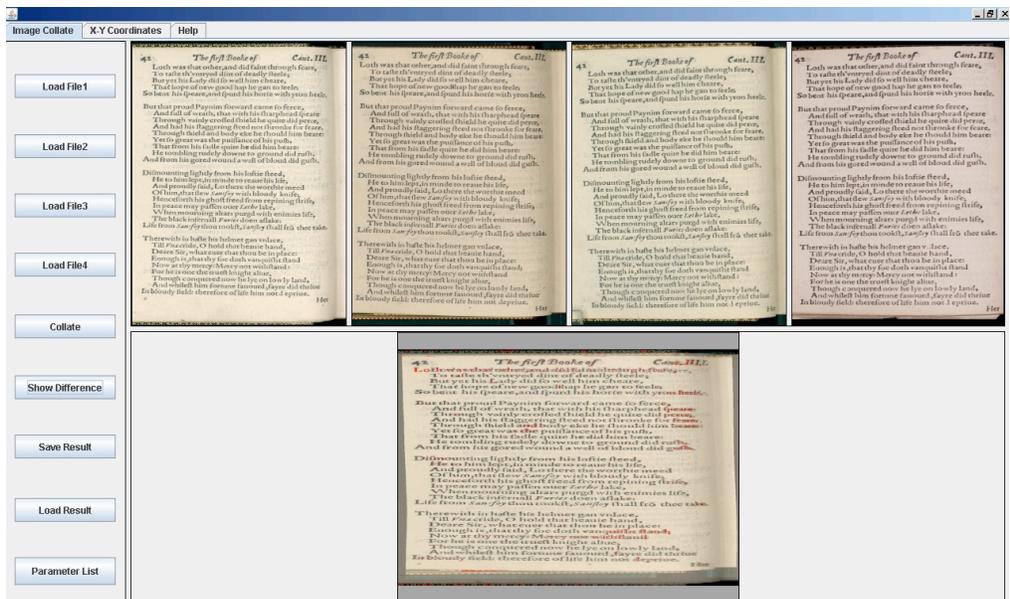


Figure 3: Differences shown as red marks

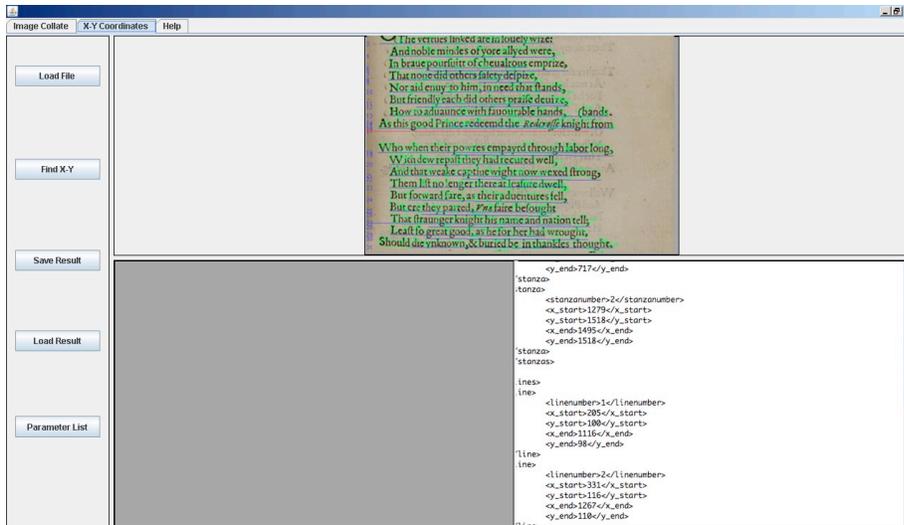


Figure 4: Finding X-Y coordinates, coordinates shown as xml format in the bottom right panel.

Collaboration can bring us more source images to refine our algorithms, targeted users to evaluate our system, and opportunities to integrate our system with other humanities projects. During our grant period, we have already attracted interest from those working in digital libraries, digital humanities, and digital preservation. We anticipate that these three communities will serve as our primary audiences with special interest from projects like the Shakespeare Quartos Archive, a digital collection of pre-1642 editions of William Shakespeare's plays. A significant portion of our collaboration deals with further prototype developments to benefit these communities. Those refinements include expanding our system to work with document images that contain colors, drawings, partial matching, watermarks, burn marks, stains, ink bleeding and other degradation factors. Heavy degradation is the major factor that causes numerous false positives. Although we will never be able to totally overcome this difficulty, we want to integrate some intelligent judging system so that with a little user interaction the precision rate can reach 100%. A further refinement of our software would also deal with increasing system speed and efficiency to handle high-resolution, large-scale data sets of historical document images via virtual interfaces. Currently, the resolution and file size slows system performance speed.

Our work on the project enables us to identify both challenges and best practices for image collation. Significant challenges are that image content often resists the ubiquitous nature of software, demanding extreme levels of customization; and that costs associated with MATLAB and other image processing software can be prohibitive to humanists. For best practices, we recommend including the following standards into document image capturing procedure:

1. Books should be captured one page at a time; when capturing, the other page should be held vertical to the binder.
2. Camera should be held perpendicular to the document. The height of the camera to the document and the distances to four corners of the document should be recorded as metadata inside the image file.

We also recommend OCR to add intermittent results, like positions of the glyphs (letters or words), to image metadata. That would help automate text-linking-image process much more easily and more accurately.

In consultation with Drs. Wang and Guiliano, lead programmer Jun Zhou is completing an article

that outlines the project, its methodology, and in-depth examination of the computational code/software designed, and the project results. Jarrell Waggoner and Jun Zhou also submitted "Image Registration for Digital Collation" as a conference paper to the International Conference on Document Analysis and Recognition (ICDAR:

<http://www.icdar2011.org/EN/volumn/home.shtml>) that will be held September 18-21, 2011 in Beijing, China. ICDAR is sponsored by the International Association for Pattern Recognition (IAPR) TC-10 (Graphics Recognition) and TC-11 (Reading Systems). The submission is currently under consideration. "Image Registration for Digital Collation" is currently available at: <http://cdh.sc.edu/resources.html>. The Sapheos prototype is available for download at [www.cdh.sc.edu/paragon](http://www.cdh.sc.edu/paragon)

Continuation of the Project: 1.) Dr. Miller and Zhou have established a partnership with the Text-Image Linking Environment (TILE) project at the Maryland Institute in the Humanities led by Doug Reside and Dave Lester. The Text-Image Linking Environment (TILE) is a web-based tool for creating and editing image-based electronic editions and digital archives of humanities texts. TILE observed a similar problem with the text-image linking process. In February 2011, the TILE project published its tool prototype. This prototype does not automatically segment image or link the text; instead it links any text block with any image block manually selected by the user and records linking in standard JSON format. In effect, it circumvents automated linking by inputting a manual identification process as central to creating the associative relations. A TILE-Sapheos partnership offers the potential for a robust creation, collation, and editing environment for digital materials that will specifically address the issue of individual character text-image linking. TILE will be experimenting with Sapheos and vice versa in the coming months in order to evaluate the best mechanisms for seamless integration; 2) Drs. Miller, Guiliano, Wang, and lead programmer Zhou are currently in the process of authoring a stage 2 project, renamed Paragon: Intelligent Digital Collation and Difference Detection. Paragon will acquire more images from national and international repositories to refine the current collation system; redesign the software architecture to speed up the collation process and to include scalable graphical interactive interfaces- standalone and on the web - for users from independent scholars to international projects; and sponsor workshops that train humanities users and provide long-term consultation services to their projects. Our final goal for Paragon is to crack the image collation nutshell, enabling multiple forms and applications of image comparison to benefit a wide range of potential users-

**The Sapheos Project:  
Transparency in Multi-image Collation, Analysis, and Representation  
White Paper**

Jarrell Waggoner, Dhaval Salvi, Jun Zhou, Song Wang, Jennifer Guiliano  
Department of Computer Science, Center for Digital Humanities  
University of South Carolina

The Sapheos Project was funded via a Level II: Digital Humanities Start Up Grant to pioneer an innovative product: a digital collation software, prototyped in MATLAB and delivered as an open-source project using C code, that provides both a back-end collation tool and a powerful front-end interface for interacting with large datasets of books. The Sapheos Project partnered with the NEH-funded Spenser Project ([spenserarchive.org](http://spenserarchive.org)), with supporting faculty from Cambridge University, Washington University at St. Louis, Pennsylvania State University, and the University of Virginia. This partnership offered the opportunity to utilize a broad cross-section of materials for testing purposes as well as an established community of scholars to advise on the context of the images being utilized with the project.

“Collation is the time-consuming but necessary comparison of two witnesses (copies) of an early modern edition in order to ascertain information about the printing process. Used extensively by historians of the book, bibliographic scholars, and those interested in the material culture of print, textual collation takes on a new importance when the underlying manuscript of the text is nonexistent. For many early modern authors whose manuscripts do not survive—Shakespeare and Spenser are perhaps the best known, but there are hundreds of lesser known authors for whom the printed text is the only known authority—textual collation is important not just for reasons of material culture and print history, but also to establish the authority of the underlying text, to separate error from accident, and isolate the way the text is from the way the text ought to be... In performing collation, researchers isolate difference as a series of binary judgments, building alteration sequentially by comparing many individual witnesses to a given ‘control copy’ that arbitrarily fixes the text to a given state. While many of the differences between texts consist chiefly of mechanical or human error—errors in typesetting and mechanical errors common to early presses and the methods for aligning and securing type within the forms—there are more than a few instances of variance within an edition that simply can’t be explained in terms of error. For example, in the 1590 *Faerie Queene* [a work of Edmund Spenser], there are variants with whole words inserted or deleted, lines abridged or added, sonnets relocated or re-ordered. Coupled with the lack of a stable underlying manuscript to fix the ground of the text, collational variance becomes part of the assemblage of the text, irreducibly part of the play of its intricate meaning.” [1]

Fundamentally, Sapheos Project software links the words on a page to the images of the words, figure tags on a page to images of the figures, and transcriptions of marginalia within the XML to images of the handwriting. Our software also automatically sections and generates (x,y) coordinate pairs for page images. Written in Java and being ported to MATLAB for C compilation, this software takes images of pages with existing XML markup and inserts (x,y) coordinate pairs into hierarchical elements—lines, paragraphs, stanzas (line groups), and figures—to allow XSL transformations to closely associate textual and image data that may vary in terms of quality and size for users. Taken together these two functionalities advance the use of

digital image processing techniques to automatically analyze historical documents with minimal user intervention. This is particularly significant for medieval and early-modern scholars who grapple with closely related sets of manuscripts that are differentiated by minute characteristics including textual variants, embellishments, and handwritings. Thus, the intent of the Sapheos Project was to create a ubiquitous tool for back end collation and user interface to allow for non-computational specialist scholars to undertake digital collation on their own. As will be outlined below, the effort to create a ubiquitous tool for digital collation was met with significant challenges:

#### **A. Image acquisition, file format and size:**

Thanks to the Spenser Archive Project, we obtained an initial dataset consisting of early modern books that had been photographed by multiple repositories (Washington University in St. Louis, Yale University, the Harry Ransom Center at the University of Texas-Austin, and others). These images appear in a variety of formats (tiff, jpg, png) and masters (uncropped, single page or double pages, with color bars and rulers present, colored or gray-scale); source image size ranged from 1MB to 128MB. Above all, to avoid distressing rare materials, the images were taken by camera held perpendicular to the books at a ninety-degree angle. This results in the common but serious problem that existing approaches have failed to resolve: warping or shading and curved text lines in the spine area of the bound volume. Image warping will not only reduce the OCR accuracy, but also impair readability. For example, figure 1 shows it is an unrealistic task to collate two similar pages by overlaying the warped page images.

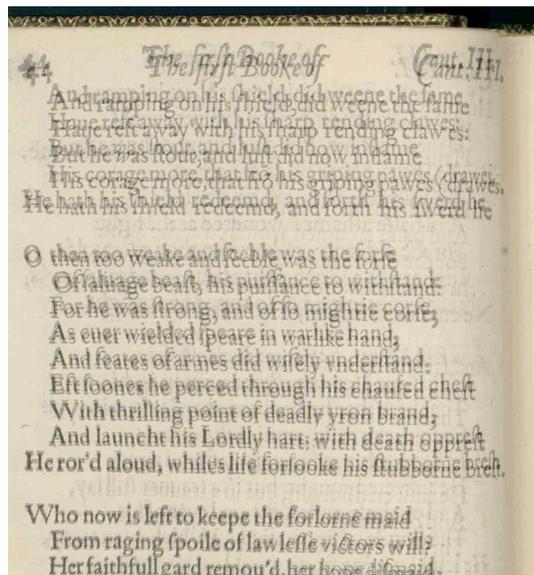


Figure 1: collation failed because of blurs caused by warping

#### **B. Image preprocessing:**

Image preprocessing included cutting the boundaries, splitting two-page images into one-page images, rough rotation and scaling, etc. We wrote several MATLAB (a high-level technical computing language) scripts to process these images in a batch order. By processing these in batch, we are able to alter multiple files at a time thereby decreasing the amount of time it takes to prepare the raw files for use.

### C. Linking document images with text:

In our initial technical plan, we suggested that we would link page images to text encoded in XML. This process involved two steps: document image segmentation and injecting the segmentation information (xy-coordinates values) into XML. Here the XML file serves not only as a container for the result, but also as a reference to segmentation procedure. For example, <hi rend="face(ornamental)"> in XML represents a large ornamental letter image that may cross multiple lines. <lg> and <l> indicate the content is a poem that should be treated as a line instead of a paragraph. We had originally intended to link every word, line, paragraph, picture in the transcript to its coordinate position in the image to enable other functions like the highlighting of an image of searched-for text. To this end we can roughly segment out words, lines and stanzas. But with unusual fonts, warping and degradation present in document images, automated segmentation can never reach a 100% precision rate.

As manual correction from mis-segmentation presented in the first step, the tagging and coordinate-mapping process took longer than anticipated. We approximated a six-week process, yet even with the tool, linking text to image proved unfeasible as it took significant labor to complete every single word and line of text. For example, see figure for page image and following segmentation result in xml format.

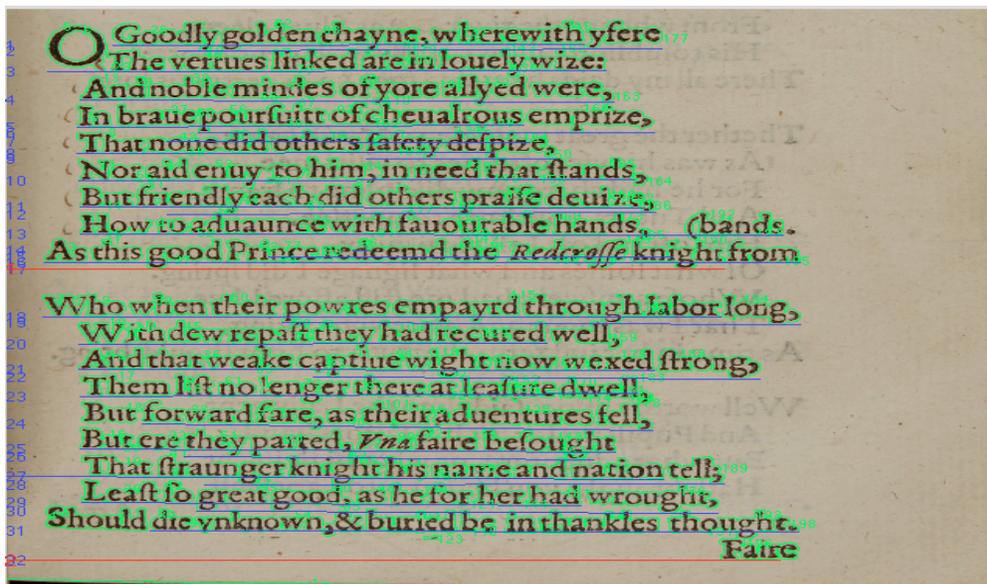


Figure 2: page with segmentation result illustrated in the image.

```
<stanzas>
:
<stanza>
  <stanzanumber>2</stanzanumber>
  <x_start>1279</x_start>
  <y_start>1518</y_start>
  <x_end>1495</x_end>
  <y_end>1518</y_end>
```

```

</stanza>
</stanzas>
<lines>
:
<line>
  <linenumber>32</linenumber>
  <x_start>1279</x_start>
  <y_start>1518</y_start>
  <x_end>1495</x_end>
  <y_end>1518</y_end>
</line>
</lines>
<words>
:
<word>
  <word>202</word>
  <x_start>1482</x_start>
  <y_start>1395</y_start>
  <x_end>1530</x_end>
  <y_end>1440</y_end>
</word>
</words>

```

This page has 2 stanzas, 19 lines and 136 words. The tool gives us 2 stanzas, 32 lines and 202 words. The manual correction for this single page needs to compare every element in the result with the image. After correction, injecting the resultant xml into the existing TEI XML transcription file requires the interjection of every TEI XML tag relating to the resulting xml element. This involves another round of manual correction if any mismatches exist. For example, if initial manual correction segments “&” as a single word, but the transcription put “&” and “buried” as one word, then that mis-identification needs to be corrected manually. As a result, we believe designing an intelligent automatic linking system is a much larger project than feasible under the auspices of this grant.

The NEH-funded TILE project (<http://mith.umd.edu/tile/>) at the Maryland Institute for Technology in the Humanities observed a similar problem and is currently developing a tool to improve the text-image linking process. In February 2011, the TILE project published its tool prototype. This prototype does not automatically segment image or link the text; instead it links any text block with any image block manually selected by the user and records linking in standard JSON format. In effect, it circumvents automated linking by inputting a manual identification process as central to creating the associative relations.

**D. Separating printed text and hand-written text:** Separating printed text and hand-written text was treated as a writer identification process [3][4]. The writer identification process is to identify writers by analyzing the structural properties of the handwriting, e.g. average height, the average width, the average slope and the average legibility of characters. In this process, the printing press was marked as a special writer. Any text content that did not display features

similar to those of the printed text are marked as hand-written text. Since computer-aided writer identification is a quite mature process in modern document image processing field, we borrowed three algorithms that worked best on our initial image dataset. They are autocorrelation (comparing shifted copies of images themselves), vertical and horizontal black run (counting number of continuous black pixels), and vertical and horizontal white run (counting number of continuous white pixels) - figure 2 shows the algorithm, and figure 3 shows the testing result. These algorithms have been implemented in the open source software – handanalyser (<http://sourceforge.net/projects/handanalyser/>).

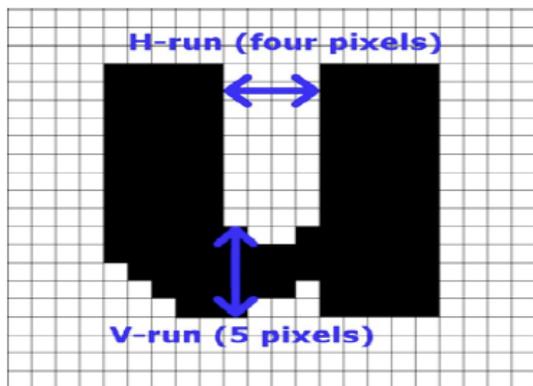
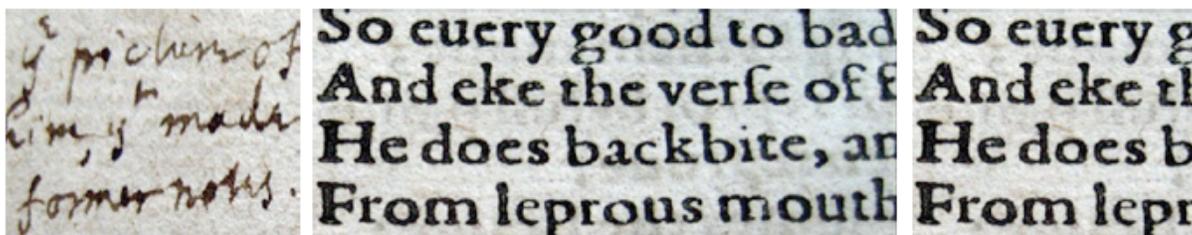


Figure 2: four continuous horizontal white pixels and five continuous vertical black pixels



A

B

C

A: Hand-written text

B: printed content

C: printed content

V-Runs	A	B	C
A			
B	39.9		
C	38.5	0.6	

D

Auto	A	B	C
A			
B	2.7		
C	2.5	0.6	

E

Figure 3: A, B and C show testing data examples; A shows a piece of hand-written text image; B and C show pieces of printed text from the same page source; D shows result running on vertical run algorithm; E shows result running on autocorrelation algorithm

**E. Automatic Image Collation:** Image collation is a process that identifies differences by comparing multiple images of similar documents. The differences range from semantic

difference to variations in the spacing or configuration of individual characters. The existing mechanisms in place for performing collation have traditionally been limited by requiring pristine scanned duplicates of the original witnesses to be compared. Three issues arise with these mechanisms: 1) they are usually obtained by flattening pages of a bound book, risking damage to the binding and adjacent pages; 2) without an efficient difference detection system, scholars sometimes need rigorous character-by-character manual inspections, e.g. typographic analysis; 3) or they require complex optical systems utilizing expensive hardware solutions. We have established an automatic image collation or automatic image registration that addresses these programmatic issues.

With two similar images, the target and the template, placed side by side, we first select about 70 to 100 pairs of matching points that are evenly distributed across the images [figure 4]; or example, the point under “Then” at the first line on the template page (left) should match the point exactly under “Then” at the first line on the target page (right). Based on these matching points, we warp the target image to the exact shape of template image [figure 5] – a process using the thin-plate-spline transformation algorithm in advanced digital image processing. Finally, we overlay the warped image and template image, resulting in the differences being shown as blurs on the overlaid image. [figure 6]

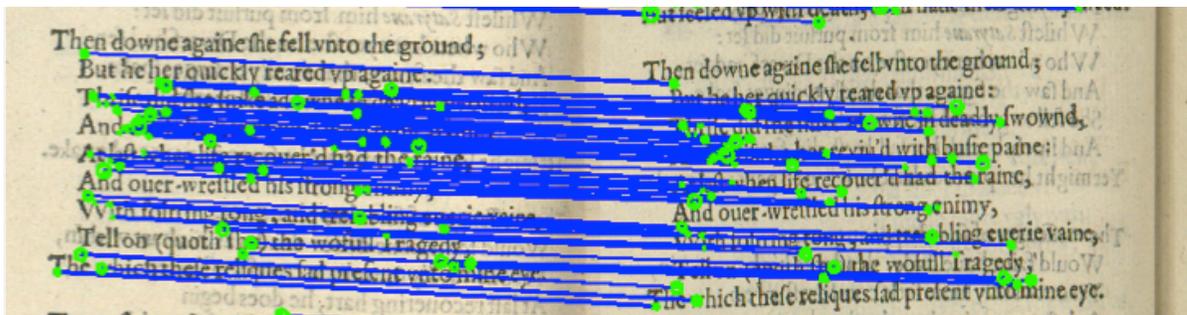


Figure 4: Point Matching

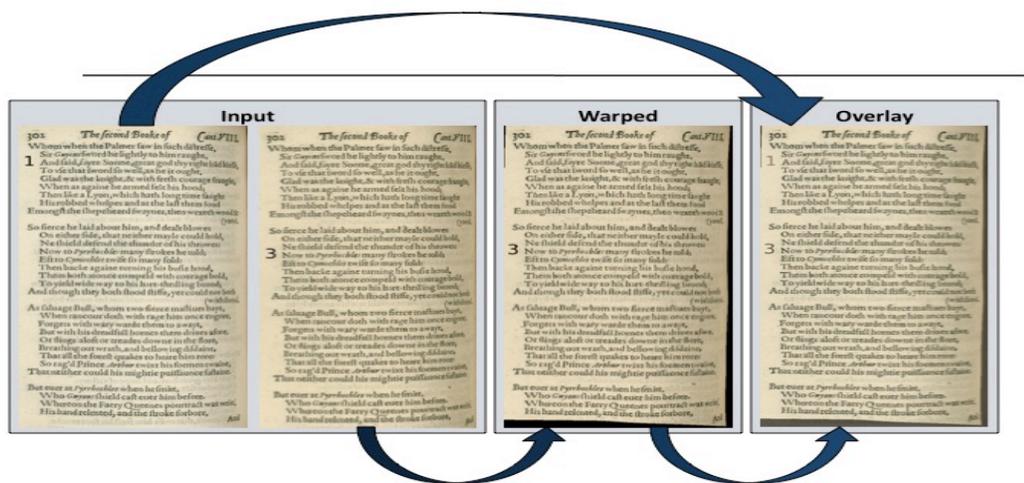


Figure 5: Warping target image to template image and overlaying them.

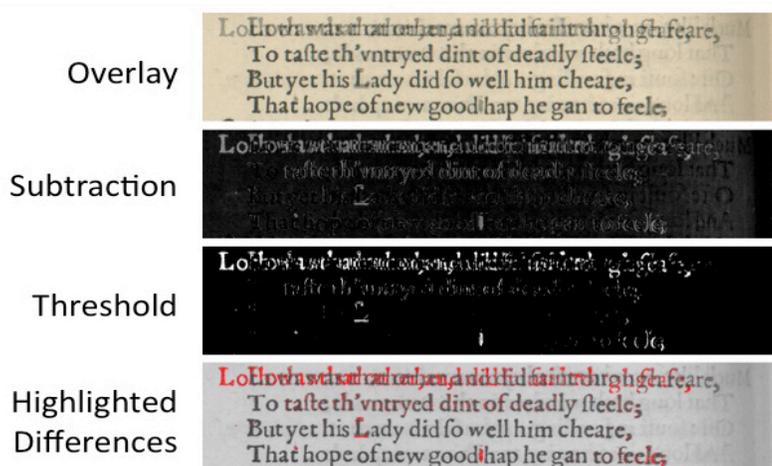


Figure 6: Difference Highlighting

To automate the collation process, precise point matching turns out to be the most critical step in the whole project. For each pair of images, we initially generate thousands of matching points using SIFT algorithm – a state-of-the-art image processing algorithm that isolates matching points. [2] We then apply a four-tier system of filters to get rid of unnecessary points.

1. Nearest neighbor match filter: good matches should happen within designated neighboring area, e.g., the point at the bottom of template page should never match the point at the top of target page.
2. Duplicate match filter: good matches should be one-to-one matches; this is required by Thin-plate-spline transformation process.

3. Thin plate spline regression filter
4. Local consistency filter

The thin plate spline regression filter and local consistency filter are advanced image processing filters that remove matches affected by noise like document degradation, ink bleeding, etc.

The prototype was implemented with MATLAB. We evaluated this prototype by selecting about 172 large size images (1279X2444) and did ten rounds of two-page collation tests. Each test round took about 12 hours running on an eight-core processor MacPro server. All of our tests showed that all differences could be indentified correctly. However, false positives were a significant challenge and are addressed below.

**F. User Interface:** Constructing a user-friendly interface that can efficiently analyze historical document images was precisely our intention, and we have succeeded in completing this endeavor. Our user interface was built as a standalone JAVA application. As a cross-platform language, JAVA turns out to be the first choice for other popular image systems, e.g. ImageJ and Virtual Light Box. As promised, our application allows user to load up to four pages. Using the choice buttons at the panel bottom, the user can select and deselect any two of them to collate [figure 6]. The pages can be made translucent in order to provide a better visual result (differences shown as red marks) [figure 7]. Most important of all, while doing collating, the user does not need to align the images by tedious dragging manually. All of these tasks are performed automatically, even with warping presented in the images. The results are displayed as two pages perfectly overlaid with difference shown as blurs. Finding x-y coordinates is also included in the user interface [figure 8]. The application would be able to print out x-y coordinates of each word and line.

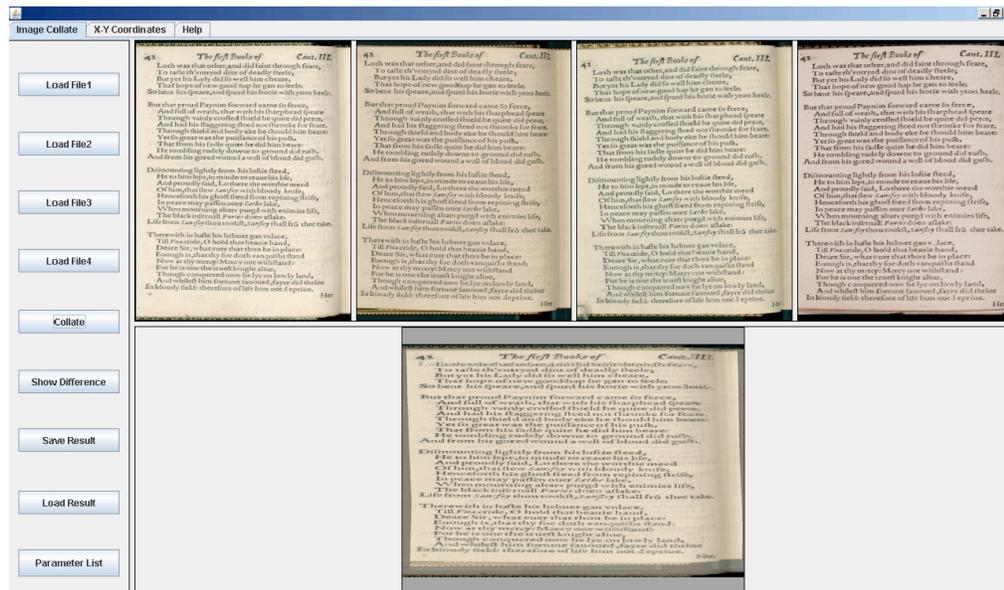


Figure 6: Collating multiple pages, result of two-pages images shown in the bottom right panel.

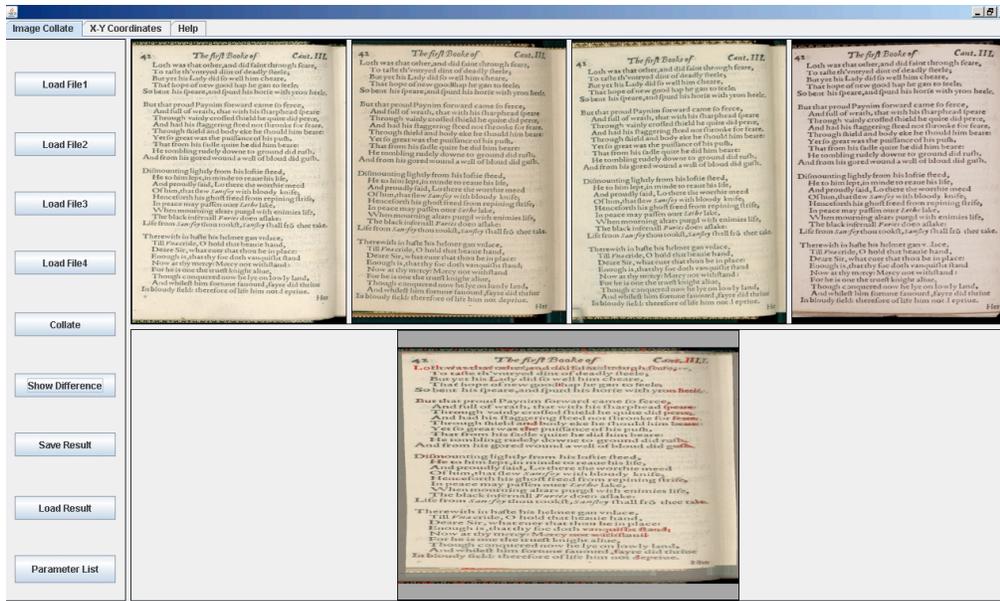


Figure 7: Differences shown as red marks

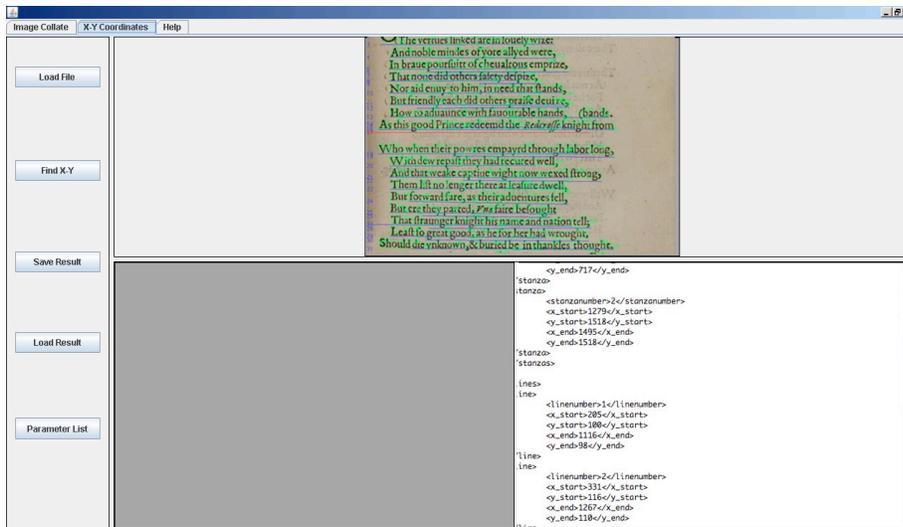


Figure 8: Finding X-Y coordinates, coordinates shown as xml format in the bottom right panel.

We had originally built an interface that would allow users to upload single images and collate two images at one time. This works perfectly for those who have small or middle scale images as it is configured to use limited computational resources common to the average desktop computer. But what about collating a significant number of large-sized books, each with hundreds of pages, in one batch? Our experience shows collating two large-sized images, say 2200 X 2400, on a 2.26G intel core and 4GB memory PC can take up to 20 minutes per image. This is not ideal, as the need to collate hundreds or even thousands of images would dramatically affect both computational performance and the amount of time that a user would need to be stationed at the computer. We intend to continue expanding our software package, so that this software can cover not only personal needs, but also these large-scale institutional projects.

Further funding will be sought to continue the project. As any application developer working today is well aware, turning a successful prototype into fully functional system needs multiple software engineering phases. We intended to develop this innovative software system into robust, open-source software that will benefit all humanists. This section outlines activities that should be involved in our continuing work:

**G. Collaboration:** Collaboration can bring us more source images to refine our algorithms, targeted users to evaluate our system, and opportunities to integrate our system to other humanities projects. During our grant period, we have begun to explore possibilities for collaboration with projects such as the Shakespeare Quartos Archive group. We intended to hold workshops with computational scientists working in the digital humanities to assess the system and provide feedback to a multi-disciplinary community of scholars engaged in digital collation – a constituency including digital collections librarians, museum staff and digital humanists. In addition, we hope to offer an initial period of high-level technical support and system adjustment for participants. Through these activities, we encourage the adaptation of this software system to preserve and analyze digital assets while acquiring more image data to refine our software system.

**H. Software refinement:** As we discussed above, turning a prototype into a robust, fully function system requires a lot of refinements. Those refinements include expanding our system to work with document images that contain colors, drawings, partial matching, watermarks, burn marks, stains, ink bleeding and other degradation factors. Heavy degradation is the major source of numerous false positives. Although we will never be able to totally overcome this difficulty, we want to integrate some intelligent judging system so that with a little user interaction the precision rate can reach 100%. A further refinement of our software would deal with increasing system speed and efficiency to handle high-resolution, large-scale data sets of historical document images. Currently, the resolution and file size slows system performance speed.

### **I. Grant products.**

In conclusion, our aim was both a) to demonstrate how historical document can be automatically collated and analyzed and b) to create prototypes that implement our algorithms.

a. A couple of papers are posted online at [www.cdh.sc.edu/paragone/resource.html](http://www.cdh.sc.edu/paragone/resource.html), including this technical white paper and “Image Registration for Digital Collation” by Jarrell Waggoner and Jun Zhou submitted to ICDAR (International Conference on Document Analysis and Recognition) 2011.

b. The prototype described here is available for download at [www.cdh.sc.edu/paragon](http://www.cdh.sc.edu/paragon).

### **J. Glossary:**

#### **Matlab**

a numerical computing environment and fourth-generation programming language.

#### **C**

a general-purpose computer programming language.

**Java**

a general-purpose, concurrent, class-based, object-oriented, cross-platform programming language.

**XML**

a set of rules for encoding documents in machine-readable form.

**Feature**

In computer vision and image processing, the feature is a piece of information that is relevant for solving the computational task related to a certain application.

**Thin-plate-spline**

the non-rigid transformation model in image alignment and shape matching.

**References:**

- [1] Randall Cream, “The Sapheos Project: Transparency in Multi-image Collation, Analysis, and Representation,” NEH Grant Application, 2009.
- [2] Lowe, David G. (1999). “Object recognition from local scale-invariant features”. Proceedings of the International Conference on Computer Vision. 2. Pp. 1150-1157
- [3] Peter A. Stokes, “Palaeography and Image-Processing: Some Solutions and Problems”, Digital Medievalist 3 (2007-2008)
- [4] Bulacu, Marius, Lambert Schomaker, and Louis Vuurpijl. 2003. Writer-identification using edge-based directional features. ICDAR, 2:937-941

**Participants:**

**Dr. Jennifer Guiliano**, PI, Associate Director of the Center for Digital Humanities at the University of South Carolina

**Dr. Song Wang**, Co-PI, Associate Professor in the Department of Computer Science at the University of South Carolina

**Jarrell Waggoner**, Ph.D Candidate in the Department of Computer Science at the University of South Carolina

**Jun Zhou**, Lead Programmer at the Center for Digital Humanities at the University of South Carolina

**Dhaval Salvi**, Ph.D Candidate in the Department of Computer Science at the University of South Carolina

**Dr. Randall Cream**, Post Doctoral Scholar in the Department of English at the University of South Carolina (2008- May 2010)

**Maliek Mcknight**, Programmer at the Center for Digital Humanities at the University of South Carolina

**Shawn Maybay**, Digital Humanity Specialist at the Center for Digital Humanities at the University of South Carolina

**Bernardo Quibiana**, Programmer at the Center for Digital Humanities at the University of South Carolina

**Citation:**

Jennifer Guiliano, Neil Fraistat, Travis Brown, Trevor Muñoz, and Seth Denbo. "Shared Horizons: Data, Biomedicine, and the Digital Humanities" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities  
Jennifer Guiliano  
Neil Fraistat  
Travis Brown  
Trevor Muñoz  
Seth Denbo

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

**Citation:**

Jennifer Guiliano, Neil Fraistat, Travis Brown, Trevor Muñoz, and Seth Denbo. "Shared Horizons: Data, Biomedicine, and the Digital Humanities" National Endowment for the Humanities, Grant Submission, University of Maryland, College Park, MD, 2011.

**Licensing:**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Collaborating Sites:**

University of Maryland  
Maryland Institute for Technology in the Humanities

**Team members:**

Maryland Institute for Technology in the Humanities  
Jennifer Guiliano  
Neil Fraistat  
Travis Brown  
Trevor Muñoz  
Seth Denbo

**Acknowledgments**

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the collaborating institutions or the National Endowment for the Humanities.

## **‘Omics’ in the Humanities<sup>1</sup>**

### **Proposal for an interdisciplinary symposium co-hosted by the National Endowment for the Humanities, the National Library of Medicine, and the Maryland Institute for Technology in the Humanities**

“Culturomics,” wrote Erez Lieberman Aiden and Jean-Baptiste Michel in their 2010 article on the quantitative analysis of culture using millions of books, “is the application of high-throughput data collection and analysis to the study of human culture.”<sup>2</sup> The highly visible efforts of the Google Ngrams team are only the most conspicuous example of the recent growth of interest in computationally-assisted scientific approaches to large-scale data and visualization in the humanities. Lieberman Aiden, who recently won a \$2.5 million New Innovator Award from the National Institutes for Health, is a rare example of a scholar with a background in genomics and biomedicine – as well as a graduate degree in history – who is demonstrating how research techniques from the medical sciences might also be used to explore topics in the humanities.

The National Endowment for the Humanities has taken a leadership role in exploring how scientists and humanists might work collaboratively toward common research goals. Grant programs like the NEH-led Digging into Data Challenge have sparked worldwide interest in how computationally driven approaches might be harnessed for humanities research. By working in conjunction with the National Science Foundation and six other international funders, the Digging into Data Challenge has sponsored teams that join humanities researchers with colleagues from the computer, information, social, and library sciences. These interdisciplinary collaborations have proven to be fruitful and led to exciting research in exploring large-scale humanities data. But as suggested by the work of Lieberman Aiden, there is another important discipline that has been largely absent from this conversation: biomedicine. Just as experts in digital humanities use computational methods to address questions in the humanities, researchers in the field of “bioinformatics” and its sister discipline “computational biology” use and develop

---

<sup>1</sup> This title is tentative and subject to discussion with ODH and NLM.

<sup>2</sup> Jean-Baptiste Michel et al., “Quantitative Analysis of Culture Using Millions of Digitized Books,” *Science* 331 (December 2010), accessed December 17, 2010, doi: 10.1126/science.1199644.

technology-based approaches to studying biology and biomedicine. Like the digital humanities, the field of bioinformatics is a “hybrid field” bringing together researchers from numerous disciplines across the academy, including biology, computer science, and engineering. As both digital humanists and bioinformaticists are developing new, computationally-based research methods, there may well be common points of interest and intersection that could be valuable to both communities.

The Maryland Institute for Technology in the Humanities (MITH), working in cooperation with the Office of Digital Humanities of the National Endowment for the Humanities and the National Library of Medicine of the National Institutes for Health, proposes a symposium to: (1) address questions about collaboration, research methodologies, and the interpretation of evidence arising from the interdisciplinary opportunities in this burgeoning area of biomedical-driven humanities scholarship; (2) to investigate the current state of the field; and (3) to facilitate future research collaborations between the humanities and biomedical sciences. The “‘Omics’ in the Humanities” symposium would include scholars working in bioinformatics, genomics, and related fields as well as researchers from the digital humanities. We aim to create opportunities for disciplinary cross-fertilization through a mix of formal and informal presentations combined with breakout sessions, all designed to promote a rich exchange of ideas about how large-scale quantitative methods can lead to new understandings of human culture.

### **Rationale and focus**

In the hard sciences the suffix ‘omics’ (pronounced like “genomics,” not “economics”) refers to a set of fields that share a focus on mapping information objects and finding interactions and relationships among those objects in large datasets. Such fields are holistic, emphasizing interconnections and inter-dependencies. Genomics, to name the best-known example of many, studies the entirety of the encoded hereditary information for a species. Much of the work in genomics and other “omics” fields is driven by researchers in bioinformatics, using the latest in computationally-based research methods. Recently, data-rich humanities research has begun to be thought of using “omics” terms, and projects have even utilized the suffix in names such as ‘Culturomics’<sup>3</sup> and the NEH-funded ‘Lexomics’.<sup>4</sup> Instead of genes and proteins that encode the

---

<sup>3</sup> “Culturomics,” <http://www.culturomics.org/>

<sup>4</sup> Wheaton College “Lexomics,” <http://wheatoncollege.edu/lexomics>

data necessary for organic life, the information objects are, for example, books, newspapers, song lyrics, and the words from which they are constructed. The scale at which these cultural objects are now available in digital form and the advances in technologies for analysis have put computational methods at the forefront of the research agenda in several humanities fields. Quantitative analysis of print culture and language that borrows and adapts scientific methodologies, especially from biology, enables new questions and innovative means to address traditional concerns.<sup>5</sup>

Although the symposium will address techniques and methodologies associated with large volumes of data, it will be more focused than just examining the use of “big data” in the humanities. First, the symposium will bring together researchers from domains that are often thought to be highly divergent in the way they create knowledge: bioinformatics and the digital humanities. Biomedical scientists are thought to concentrate exclusively on experimental aspects of knowledge production driven by “pure data” approaches while humanities scholars are often seen as being driven by opinion-based analysis that uses data to prove a point rather than as an argument in itself. Second, by focusing on ‘-omics’ in the humanities the symposium will look at specific modes of analysis rather than those that are merely rich in data. As distinguished from other uses of scientific methodologies for cultural analysis, the ‘-omics’ approach to knowledge in the humanities potentially leads to results providing broad views and allows questions about long-term change in cultural, historical, and linguistic meaning. Sequence alignment and network analysis will serve as our particular focus because they already offer models of fruitful existing collaborations that will enrich our effort to explore the intersections of computational biology/bioinformatics and humanities.

In bioinformatics, sequence alignment is a primary technique for understanding the functional and evolutionary purposes of proteins. Scientists compare the relationship between parallel sequences within long strings of text that represent those proteins. Where two sequences

---

<sup>5</sup> A few examples of work using methodologies adapted from biological sciences include: M. L. Jockers, D. M. Witten, and C. S. Criddle, “Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification,” *Literary and Linguistic Computing* 23, 4 (September 29, 2008): 465-91; Matthew Spencer et al., “How Reliable is a Stemma? An Analysis of Chaucer’s Miller’s Tale,” *Literary and Linguistic Computing* 18, 4 (November 1, 2003): 407-22; and see Alberto Piazza’s (Professor Human Genetics at Turin University Medical School) afterword to Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (Verso, 2005).

share a common ancestor, differences can indicate mutations and provide evidence of the role specific sequences play in evolution. However, sequence alignment isn't only a domain of biomedicine. Digital humanities scholars working in the fields of language and linguistics have been using gene-sequencing techniques to understand how manuscripts and other literary texts change over time.<sup>6</sup> Similarly, genomics scholars have explored linguistic approaches to genetic data. Sungchul Ji, in "The Linguistics of DNA," argues that cells have a language and that "linguistics provides a fundamental principle to account for the structure and function of the cell."<sup>7</sup>

Similarly, network analysis presents collaborative potential. Scientifically-driven network analysis views information ecologically, searching for and analyzing patterns and connections across datasets. In genetics, network analysis is used to explore gene function by evaluating the relationships between and among genes with visualizations of data showing complex systems. Humanists have also had a long history of doing network analysis, studying how information, ideas, and philosophies are transmitted across time and cultures. Numerous digital humanities projects have explored how computationally-based techniques including data-mining and visualization, might be used for network analysis. In fact, a 2010 NEH summer institute that brought together digital humanists with mathematicians to explore methods for network analysis proved to be extremely valuable to all the attendees, with both disciplines learning from one another.

The Omics Symposium seeks to do something similar by bringing together researchers from the digital humanities and bioinformatics to see how they might learn from one another and, possibly, work together on future projects. As this symposium is meant only to be a brief, two-day introduction, our goals are to keep a fairly tight focus on these two research methods of common interest (sequence alignment and network analysis). Future workshops, if appropriate, could expand this focus, of course. But focusing on two methods that have uptake in both domains seems like a natural place to begin the conversation. Symposium participants will: 1)

---

<sup>6</sup> Digital Humanities journals have been one venue for articles on the application of methods from phylogenetics to textual transmission of medieval manuscripts. See H. F. Windram et al., "Dante's Monarchia as a test case for the use of phylogenetic methods in stemmatic analysis," *Literary and Linguistic Computing* 23, 4 (December 2008): 443-463; Heather F. Windram, Christopher J. Howe, and Matthew Spencer, "The Identification of Exemplar Change in the Wife of Bath's Prologue Using the Maximum Chi-Squared Method," *Literary and Linguistic Computing* 20, 2 (June 2005): 189-204; M. Spencer, et al., "Analysing the order of items in manuscripts of the Canterbury tales" *Computers in the Humanities* 37 (2003): 97-109.

<sup>7</sup> Sungchul Ji, "The Linguistics of DNA: Words, Sentences, Grammar, Phonetics, and Semantics," *Annals of the New York Academy of Sciences* 870, 1 (May 1999): 411-417.

communicate disciplinary and interdisciplinary approaches; 2) highlight the challenges of collaboration; 2) demonstrate shared understandings of data and its uses, including the interpretation of data; 3) identify opportunities for disciplinary cross-fertilization of approach and theory; 4) reconcile domain specific research processes; 4) illuminate how to evaluate and interpret evidence; and 5) suggest new avenues of exploration between humanists and scientists working in these areas. More simply, this symposium offers the opportunity to explore and solidify on-going connections between humanists and scientists and to introduce these approaches and challenges to new scholars. The application of high-throughput data collection and analysis, then, could potentially lead to new ways of understanding both human culture and the natural sciences.

### **Cooperation Between NEH, NLM, and MITH**

Just as the Omics in the Humanities symposium represents a unique opportunity to bring together researchers from two distinct areas of scholarship – the humanities and biomedicine – it also brings together two of the largest funders of such research in the United States. This cooperation, between the National Endowment for the Humanities’ Office of Digital Humanities (ODH) and the National Institutes of Health’s National Library of Medicine (NLM), is a critical part of the Omics symposium. The symposium is very much a first step towards building a new bridge between two research communities – the digital humanities and bioinformatics. Both ODH and NLM are committed to working with the University of Maryland on the symposium agenda, publicity, and outputs to ensure it is useful to their respective communities.

The ODH, as one of the leading funders of digital humanities, is well-positioned to help ensure that the right scholars attend the meeting to make it as productive as possible. Similarly, the NLM, part of the NIH – the largest funder of biomedical research in the world – will help to ensure that influential people from the biomedical community with an interest in bioinformatics and interdisciplinary research are in attendance. Also, staff from the ODH (and from other units of NEH) as well as staff from NLM (and other units of NIH) will participate in the symposium with an eye toward identifying possible funding opportunities for researchers working at the nexus of the humanities and biomedicine. With its role as the world’s largest medical library, the NLM also brings to the symposium a long history of facilitating research projects involving both

scientists from the biomedical community and humanists from the history of medicine and science and related fields.

MITH and the University of Maryland are well-positioned to organize and host this event. MITH offers expertise in digital humanities, natural language processing, and data curation while the University of Maryland supplies deep expertise in biomedicine, bioinformatics, linguistics, and information science. We have assembled an advisory board of scholars from various humanities departments, the iSchool, the Center for Bioinformatics and Computational Biology, the Institute for Advanced Computer Studies, and the Department of Computer Science to plan and participate in the event. With the help of this board and the advice of ODH and NLM, we aim to draw top international scholars as participants and speakers. Holding the event in the Washington, DC area will facilitate participation by staff of both NEH and NIH. We have every expectation that this event will prove as successful as the 2007 “Summit Meeting of Digital Humanities Centers and Funders,” which MITH similarly co-hosted (via a cooperative agreement) with NEH.

### **Symposium Participants**

We anticipate fifty people attending the symposium representing a diverse cross-section of disciplines. Approximately half of those will be participants (speakers, chairs, respondents, and advisory board members). General attendance will be open and places will be allocated by reservation. Participants for the symposium will be a mixture of invited (40%) and open-call (20%) attendees with 20% being reserved for organizers and NEH attendees. Open-call participants will be selected based on a widely-distributed call that will request a c.v., a statement of research interest, and an 500-1000 word abstract for a 15 minute paper that would address the use of either sequence alignment or network analysis within the attendees’ research. Selected papers would be distributed to symposium attendees. Open-call participants will be selected by the Advisory Board in consultation with NEH and NLM based on the following criteria: 1) scholarly engagement with sequence alignment and/or network analysis; 2) quality of proposed paper; and 3) collaborative potential. Invited attendees will be selected by the Advisory Board in consultation with NEH; a preliminary list of candidates for consideration is provided below.

### **Symposium Format**

The symposium will be a two-day event preceded by a public keynote lecture showcasing the event itself and raising public interest in the key intersections of science and humanities research. Jeremy John, Curator of eManuscripts at the British Library, or David Searls, Professor of Genetics and former Vice-President for Genomics at GlaxoSmithKline, are potential candidates and would each make an ideal introductory keynote speaker for the symposium. John is an evolutionary biologist whose primary research interests, including phylogenetics, bioinformatics, manuscripts, and the history of computing unite in his use of evolutionary theory to study digital preservation and digital life stories. Searls, a geneticist, explores culture, language, and bioinformatics, most notably in an article in the *Journal of Computational Biology* titled “From Jabberwocky to Genome: Lewis Carroll and Computational Biology.” Either of these potential keynoters would offer the ability to speak to both humanists and scientists, with their core research uniting these disciplines into a cohesive scholarly framework.

We propose that the first day begin with a session on the history and theory of the crossover between digital humanities and bioinformatics. This will encourage a common understanding of the origins and philosophy that shape our intended goals while simultaneously creating a shared lexicon from which participants can draw. Following this theoretical grounding, the first day will focus primarily on sequence alignment. We will have two or three formal 15-20 minute presentations by scholars who submitted papers during the open-call process regarding sequence alignment. These talks will be focused on the domain-based research questions being investigated and the methodological issues faced in collaborative cross-domain work. This initial session will be followed by a roundtable discussion about the problems and challenges associated with emergent approaches in sequence alignment. Invited speakers will then give talks on the theory, methodology, and future directions for sequence alignment research. Topics for the breakout sessions will be proposed and agreed upon by the participants at close of the invited speaker presentations. This will enable the group to benefit from the collective knowledge and interests of the participants. The final plenary session of the day will be a roundtable to discuss and assess the conclusions of the breakout session, culminating in a set of “future directions” for sequence alignment. By the conclusion of day one, then, participants will have addressed issues related to the benefits and problems of scientific approaches to understanding textual change and the specifics of particular alignment techniques. They will have critiqued the kinds of questions that can (and cannot) be answered utilizing these techniques and constructed recommendations for further study and/or cross-fertilization of sequence alignment. The potential for a

conversation between biologists and humanists on the life of texts and the language of cells could be a fascinating product of the symposium, fostering conversations that could lead to radically different understandings of knowledge in both domains.

The focus of the second day would be on network analysis, particularly on questions regarding: (1) commonalities between social and biological networks; (2) types of data suitable for network analysis; and (3) the quality of data and the accuracy of network modeling. Participants will explore the problem of divergent approaches to hypothesis formulation and data-driven research to expose the distinctiveness of research aims across types of network analysis. This will promote new ways of thinking about network theory and visualization. The structure of the second day will be similar to the first, including invited papers, a roundtable discussion and breakout sessions. The final session of the last day will aim to be a summation of the discussions and presentations from the entire event, and finish with action points for future activities and collaborations. Lunch on the first day would include lightning rounds featuring related work by attendees who are not themselves giving papers, and on the second day there would be a matchmaking session to encourage interaction and to provide additional networking opportunities.

**Program:**

Evening before symposium      Public keynote lecture followed by welcome reception

***Day 1: Sequence Alignment***

9:00-9:15	Welcome and overview of desired outcomes
9:15-10:00	Session 1: History and theory of the cross-over
10:00-11:00	Session 2: Research Paper Presentations - sequence alignment (CFP)
11:00-11:30	Break
11:30-12:15	Session 3: Roundtable Response to Existing Research
12:15-1:15	Lunch (with lightning rounds for attendees presentations)
1:15-2:00	Invited Talk 1 and Discussion
2:00-2:45	Invited Talk 2 and Discussion
2:45-3:15	Break
3:15-4:15	Breakout Sessions (problems, issues, developments)
4:15-5:30	Round Table (synthesizing the breakout sessions and discussing future directions)
6:30	Dinner

***Day 2: Network Analysis***

9:00-9:15      Review and introduction to Day 2

9:15-10:15	Session 1: Research Paper Presentations - network analysis (CFP)
10:15-11:15	Session 2: Roundtable Response to Existing Research
11:15-11:45	Break
11:45-12:30	Invited Talk 1 and Discussion
12:30-1:30	Lunch (with a matchmaking session for participants and attendees)
1:30-2:15	Invited Talk 2 and Discussion
2:15-2:45	Break
2:45-3:45	Breakout Sessions (problems, issues, developments)
3:45-4:45	Round Table (synthesizing and discussing future directions)
4:45-5:30	Lessons learned and future directions

### **Proposed Outcomes**

Symposium findings will be disseminated in both traditional and innovative ways. The primary published outcome will be a report on the symposium content, learned lessons, and recommendations for future activities. The report will outline both short and long-term goals that can be undertaken by humanists and scientists. It will be published via a symposium website that will be used for publicity and outreach prior to the event and to maintain a constant web-based presence afterwards. We will record all invited talks, disseminate papers, and document workshop activities to produce a robust archive of the symposium and the broader context in which the participants are working. To encourage ongoing interaction between symposium participants we will make use of existing social networking and web-based collaboration tools for sharing information and knowledge (e.g. Google Docs, Twitter). To obtain maximum value from the event, MITH would also work with the ODH and NLM to develop evaluative instruments for the symposium, potentially including surveys, content assessment tools, and interviews during the course of the symposium.

Our ambition is to achieve ongoing energetic discussion and collaboration among participants and the wider audience of humanists and scientists in order to generate a robust community working on sequence alignment and network analysis with an attention to culture. In serving not just invited participants but also the wider community through the open call and website, the symposium should achieve the objective to engage in collaboration with colleagues from diverse backgrounds. We will pay special attention to disseminating all calls, notices, and products to the various disciplines engaged in sequence alignment and network analysis activities including, but not limited to, genomics, bioinformatics, history, literature, English,

computational linguistics, and computer science.

### **Symposium Advisory Board**

MITH and the University of Maryland have formed an advisory board for symposium planning purposes. The board includes: Neil Fraistat (Department of English), Trevor Muñoz (University of Maryland Libraries), Travis Brown, Jennifer Guiliano, and Seth Denbo (MITH), Kari Kraus (iSchool and Department of English), Carl Kingsford (Computer Science & the Center for Bioinformatics and Computational Biology), Doug Oard (iSchool and University of Maryland Institute for Advanced Computer Studies), and Mihai Pop (Computer Science & the Center for Bioinformatics and Computational Biology).

Additionally, we will solicit additional advisory board members to aid us in selecting speakers including, but not limited to: Lev Manovich, Professor, Visual Arts Department, University of California-San Diego; Marti Hearst, Professor, School of Information, University of California-Berkeley; Colin Allen, Professor, Department of History and Philosophy of Science and Program in Cognitive Science, Indiana University; Jon Kleinberg, Tisch University Professor, Department of Computer Science, Cornell University; Gregory Crane, Professor, Department of Computer Science, Tufts University; Erhard Hinrichs, Professor, General and Computational Linguistics, Eberhard-Karls-Universität Tübingen, Seminar für Sprachwissenschaft, Tübingen, Germany; and Tim Hitchcock, Professor, Department of History, University of Hertfordshire, UK.

## **Potential speakers/participants**

We are hoping to invite participants who will mesh in productive future collaborations. The projects and scholars listed here pursue research involving network analysis and sequence alignment. Many are already, or have been involved in interdisciplinary projects that have research teams comprised of scientists and humanists.

**Erez Lieberman Aiden**          Harvard University, Society of Fellows

**Jean-Baptiste Michel**          Harvard University, Department of Psychology

Dr. Lieberman Aiden and Dr. Michel are the two scholars most closely associated with the Culturomics project at Harvard University and were responsible for producing the Google Ngram viewer. To date, this is the highest profile use of scientific methodologies for asking cultural questions.

**Mark Algee-Hewitt**          McGill University, Department of English

Dr. Algee-Hewitt currently serves as a post-doctoral researcher exploring the potential today's digital resources have to provide a new understanding of the socio-cultural, political and literary transformations that took place during the eighteenth century. His research builds on existing work on the history of the sublime in eighteenth-century literature through an analysis of word clusters.

**Cathy Blake**                  University of Illinois Urbana-Champaign, School of Library and Information Science; Department of Computer Science and Medical Information Science; Fellow, Environmental Change Institute

Dr. Blake's scholarly interests are in how to "accelerate scientific discovery by synthesizing evidence from text. Her techniques embrace both automated and human approaches that are required to resolve contradictions and redundancies that are inevitable in the information intensive world in which we live." <http://people.lis.illinois.edu/~clblake/>

**Katy Börner**                  Indiana University Bloomington, Victor H. Yngve Professor of Information Science

Dr. Börner, founding director of the Cyberinfrastructure for Network Science Center, provides investigations into the intersections of Network Science, Scientometrics, Knowledge Management, and Information Visualization.

**Dan Brown**                  Waterloo University, School of Computer Science; Bioinformatics

## Research Group

Dr. Brown has done work on automatic detection of rap music rhymes and has also written on using computers to determine the correct phrase when someone searches for a misheard song lyric. The methods have also been applied to predict scansion of poems and to characterize rappers by their rhyme structure. There is also a relevant project in the department of Bioinformatics at Waterloo University that is about measuring distance between two information objects called “Alignment Free Whole Genome Phylogeny,” one publication of which studies chain letters to ‘show how to infer the family tree of anything that evolves, from genes to languages to plagiarized schoolwork.’<sup>8</sup>

<http://www.cs.uwaterloo.ca/~browndg/>

**Mark Davies** Brigham Young University, Department of Linguistics;

Dr. Davies primary areas of research are corpus linguistics, the design and optimization of linguistic databases, language change and genre-based variation, and frequency and collocational analyses (all for English, Spanish, and Portuguese).

**Niles Eldredge** American Museum of Natural History

Mr. Eldredge is a paleontologist who has written on using evolutionary ideas to trace the development of musical instruments. He is another potential keynote speaker. <http://www.nileseldredge.com>

**Michael Gleicher** University of Wisconsin, Department of Computer Science

Dr. Gleicher has a strong interest in computer graphics and visualization in a wide variety of areas. Current collaborations include Educational Science (comparing epistemic frames), Genetics (comparing whole genome alignments), Structural biology (protein shapes and motions), Literary Scholarship (statistical analysis of text corpora), and Virology (understanding virus evolution).

<http://pages.cs.wisc.edu/~gleicher/>

**Bernardo A. Huberman** Senior Hewlett Packard Fellow; Director of the Social Computing Research Group at Hewlett Packard Labs

Mr. Huberman focuses on “methods for harvesting the collective intelligence of groups of people in order to realize greater value from the interaction between users and information.” Huberman is one of the founders of the field of ecology of computation, and recently published *The Laws of the Web: Patterns in the Ecology of Information* (MIT Press). Huberman’s present work centers on the design

---

<sup>8</sup> Charles H. Bennett, Ming Li, Bin Ma, “Chain Letters and Evolutionary Histories”, *Scientific American* 76-81. June 2003.

of novel mechanisms for discovering and aggregating information in distributed systems as well as understanding the dynamics of information in large networks.

<http://www.hpl.hp.com/research/idl/people/huberman/>

**Jeremy John** British Library, Curator of Scientific Manuscripts

Dr. John has interests in the adaptive information technology, history of computer use; methods of measurement and analysis, influencing theory and philosophy; unpublished ideas and data; scientific exploration; evolutionary science, and informatic analysis of academic subject partitioning.

[http://www.bl.uk/researchregister/1.10/?app\\_cd=RR&page\\_cd=RESEARCHER&l\\_researcher\\_id=120](http://www.bl.uk/researchregister/1.10/?app_cd=RR&page_cd=RESEARCHER&l_researcher_id=120)

**Mark D. LeBlanc** Wheaton College, Department of Computer Science; Wheaton Genomics Research Group

Dr. LeBlanc is a primary investigator on the ‘Lexomics’ project that “hybridizes traditional humanistic approaches to textual scholarship, such as source study and the analysis of style, with advanced computational and statistical comparative methods, allowing scholars ‘deep access’ to digitized texts and textual corpora.” <http://wheatoncollege.edu/faculty/profiles/mark-d-leblanc/>

**Peter Leonard** The University of Chicago, Humanities Research Computing

Dr. Leonard is primary investigator of the Automated Literary Analysis of the Scandanavian Corpus in Google Books, a Google Digital Humanities Research Award project.

**Mark Liberman** University of Pennsylvania, Department of Linguistics, Department of Computer Science; Director, Linguistic Data Consortium

Dr. Liberman explores the gestural, prosodic, morphological and syntactic ways of marking focus, and their use in discourse as just one of his recent research areas within linguistics.

<http://www.ling.upenn.edu/~myl/>

**Aditi Muralidharan** University of California Berkeley, Department of Computer Science

Ms. Muralidharan serves as primary investigator of an NEH-funded effort to provide an open source tool for the exploration and visualization of the grammatical structure of text. She is currently a fourth year Ph.D. student at Berkeley.

**Robert K. Nelson** Director, Digital Scholarship Lab, University of Richmond

Dr. Nelson is an historian of nineteenth-century America whose professional interests encompass U.S. history through the lens of nineteenth-century race and confederate nationalism. He currently leads a

number of projects exploring the Civil War through network analysis.

**Robert J. O'Hara**                      Academic naturalist and evolutionary biologist

Dr. O'Hara's professional interests lie in systematics, biodiversity, and the history and theory of the historical sciences. His publications on the use of scientific methodologies for the analysis of text including, include Robinson, Peter M.W., & Robert J. O'Hara, "Cladistic analysis of an Old Norse manuscript tradition", *Research in Humanities Computing* 4, 115–137. 1996. <http://rjohara.net/>

**Christopher Potts**                      Stanford University, Department of Linguistics

Dr. Potts has an interest in Data-rich humanities research and data-mining. Of particular interest is his recent participation in a grant project called 'Tools for Data-Rich Humanities Research'.

<http://www.stanford.edu/~cgpotts/>

**David B. Searls**                      University of Pennsylvania, Department of Genetics

Mr. Searls is formerly Vice President for Genomics at GlaxoSmithKline. He is a computational biologist with a broad interest in the philosophy that underpins such work and has written on culture, language, and the relationship of both to bioinformatics.

<http://www.med.upenn.edu/apps/faculty/index.php/g306/c425/p6363>

**Timothy Tangherlini**                      University of California-Los Angeles, Professor, Scandinavian Section, Department of Asian Languages and Cultures; the Center for Medieval and Renaissance Studies, the Religious Studies Program, and the Center for Korean Studies and the Center for European and Eurasian Studies

Dr. Tangherlini has published widely on folklore, literature, film and critical geography. His main theoretical areas of interest are folk narrative, legend, popular culture, and critical geography. His main geographic areas of interest are the Nordic region (particularly Denmark and Iceland), the United States, and Korea. He recently served as the leader of NEH-funded summer institute on "Network Analysis for the Humanities."

**Robin Valenza**                      University of Wisconsin, Department of English

Dr. Valenza's primary interests lie in eighteenth- and nineteenth-century literature and culture through the use of large-scale full-text digital archives and tools for analyzing and classifying large amounts of "dirty" data (from over 100,000 books) alongside more traditional modes of close reading.

**Matthew Wilkens**                      Notre Dame University, Department of English

Dr. Wilkens works on contemporary literary and cultural production with particular emphasis on the development of the novel after World War II. His digital projects range from mapping the literary landscape of the American nineteenth century to identifying patterns of allegorical writing across several centuries to evaluating the convergence of international style in the age of globalization. He currently serves as vice president of the Digital Americanists society.