

**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.