

**Innovation Prospectus**  
**Fall 2009**

**Research in the Digital Humanities at the University of South Carolina**  
**Center for Digital Humanities**  
**Randall Cream, Associate Director**  
**[cdh.sc.edu](http://cdh.sc.edu)**

## Contents

Introduction & Prospectus ..... \$

### Narratives:

Humanities Gaming Institute ..... 3-13

The Sapheos Project ..... 14-20

Human Voices Project ..... 21-31

The Stibos Project ..... 32-38

Handheld Art ..... 39-54

**Research & Innovation Prospectus**  
**Randall Cream, Associate Director**  
**Center for Digital Humanities, University of South Carolina**  
October 2009

Since its inception in 2008, I have been the sole specialist in digital humanities at the Center for Digital Humanities at the University of South Carolina. I've been fortunate to benefit from a free hand to pursue innovative research and an indulgent and very hands-off faculty sponsor. Since its founding in Summer of 2008, the work of the CDH—and my own work—blossomed. We've done a lot with a very small staff of four—along with my work in Digital Humanities, we benefitted from a staff programmer and three undergraduates: one in visual communications and two in computer science.

During the academic year 2008/2009, the Center and I were responsible (PI or Co-PI) for \$4 million in proposed research, generating and supporting 10 grant proposals. In Digital Humanities, research is a necessarily collaborative affair; practitioners who labor alone are likely to repeat earlier similar work performed elsewhere and unlikely to effectively circulate their work within the field. Peer review is best conducted blind, where the best ideas rise to the top and a discipline embraces innovative ideas and personalities. But the vagaries of funding decision do not always reflect a research project's merit; we can also predict that the presence of numerous well-qualified collaborators indicate some prospect of viability for research. In the last academic cycle (AY 2008-09), the research that I conducted at the CDH enjoyed more than 40 collaborative relationships involving 27 different collaborators (more than a few multiple participants) in 3 countries working at 17 different institutions: USC, WashU, Penn State, UVa, Cambridge, McGill, UQaM, UCSB, NCSA, UIUC, Ohio State, USF, UC Riverside, and the private companies JSTOR, Google, and Collexis, as well as exemplary repositories such as Houghton (Harvard), Beinecke and Elizabethan Club (Yale), Folger Shakespeare Library, Harry Ransom Center at UTexas, and several libraries and repositories at Oxford and Cambridge. In short, the CDH at South Carolina, and my work within the CDH, has been fortunate to find ample and august collaborations with institutions and individuals by attending to pressing problems in the humanities that are of clear interest to the field.

Even while acknowledging the enormous necessity of collaboration, I take responsibility for five projects in last year's competitions: I conceived of the projects, planned the approach, communicated my vision to partners, wrote all of the grant documents, and recruited collaborators on these five proposed grant projects during the academic year 2008-09. Their failures and successes are my own. The five projects, successful and unsuccessful, are included in this document—the verbatim grant narratives are included, prefaced by a one-page description and assessment. Of the five proposals planned and written by me, two were funded and two were not funded. A fifth, the Human Voices Digging into Data project, remains under review. The successful grants involve three different departments at my home institution and 7 different USC Faculty members. Our budget at the Center for Digital Humanities last year was \$89,000. Our projects were awarded ~\$450,000, with \$300,000 pending.

I offer these pages as both a report of the work conducted last year and as a sample of my vision of scholarship in this field. Conveniently, it also serves as a writing sample. While these pages are my writing, my thinking, and my work, I cannot take sole credit for the accomplishment of that work. Anything worth doing is necessarily greater than oneself, and these projects outline a deeply collaborative approach to research. More than just innovative project designs, these narratives represent my understanding of my place in this new field, and I am both proud and fortunate to have found willing collaborators.

## Humanities Gaming Institute: Serious Games for Research and Pedagogy

**Competition: NEH Institutes for Advanced Topics in the Digital Humanities**

**Submitted: February 18, 2009**

**Awarded: \$250,000 May 29, 2009**

**humanitiesgaming.org**

Duncan Buell, Project Director

Randall Cream, Co-Director

The IATDH competition represented a particularly attractive grant category for 2009. In its second year of existence, this competition is modeled on the Summer Institutes and Seminars, but the budget amount is much larger: up to 250K in one year. In the late Fall, I worked with Duncan Buell, chair of Computer Science, to discuss an IATDH built around a data mining institute. As February approached, I began to consider gaming as a better topic for our competition. In the last days of January, I swiftly built a strong team: Duncan Buell, Simon Tarr, a Game developer from Media Art, and Heidi Cooley, a New Media theorist. I wrote the grant narrative, Heidi and I developed a reading list, and Simon recruited three highly visible experts to bring instant recognition.

Essentially, the HGI brings together an array of outside experts with the local team (Duncan, Heidi, Simon, and me) to direct an intensive three-week seminar on gaming: from theory to practice, with development workshops in the afternoons and open lab times in the evenings during the institute. We'll invite 20 participants from across the nation, looking for innovative ideas, under-represented constituencies, and emerging graduate students. The institute carries with it a year of support for the participants, including server space for online multi-player games.

The IATDH competition funded only five proposals in 2009: UCLA, UC Irvine, UVa, George Mason, and our Humanities Gaming Institute.

## **Humanities Gaming Institute: Serious Games for Research and Pedagogy**

In its ability to accurately model decision-making in a variety of adaptable circumstances, game play has emerged as a productive area of research and teaching for humanities scholars from an array of disciplines. While the utility of game-based learning and testing is acknowledged throughout primary and secondary education, it remains something of a niche player within the university. This is yet another area where higher education has resisted change, much to its detriment. Even where gaming is valued as a social networking platform, in familiar spaces such as Second Life, World of Warcraft, and Everquest, its effectiveness as a research platform remains largely untapped. We propose to host a Humanities Gaming Institute for faculty and advanced graduate students in May 2010 in order to foster a productive relationship between serious games for learning and serious games for research for faculty without access to the infrastructure necessary to undertake such work. Bringing together a collaborative group of twenty teachers, scholars, and researchers to develop serious games will help to highlight the potential in games as a platform for substantial research in the humanities and facilitate its transmission into broader venues. Our model, based on an intensive three week institute followed by a year of sustained technical support, collaboration, and infrastructural assistance, is strategically planned in order to allow our participants to not just develop and share their own research games, but we also want to assist their efforts to share their models with a broader community of scholars at their home institutions. By building into the structure of our project an ongoing infrastructure for supporting participants as they continue to re-think, develop, and re-tool their use of serious games, we can be assured of making a meaningful contribution to the sustained presence of serious games as a component of digital humanities.

We propose a three-week summer institute for twenty faculty and advanced graduate students with an interest in the application of games to research and teaching in the humanities. Convened by researchers working in the departments of Computer Science, English, Digital Humanities, Media Arts, and Film, the Institute will enjoy an interdisciplinary approach to the academic applications for serious games through the background approaches of its faculty directors. With strong support from the Lead Developer at the University's Center for Digital Humanities; two graduate students working in game development, new media, and digital humanities; and two advanced undergraduate students from the department of Computer Science, the Institute will be able to offer an extraordinary level of technical support, game planning, and game development assistance to its participants. With guest lectures by three nationally recognized leaders in the field of serious games as academic work, we can help our participants quickly move to work at the leading edge of gaming and game theory. We believe the time is right to move game-building from a specialization held by a few in academia into an environment where humanists working in a variety of fields can easily develop serious games for teaching and research.

## Significance

By adopting serious games into their research and teaching, humanities scholars are poised to deliver remarkable results in a variety of fields far beyond the obviously connected areas of media arts and digital humanities. The cultural impact of games in the lives of students is becoming so unmistakable that faculty leaders like Ian Bogost and Tracy Fullerton are increasingly blurring the line between the production and interpretation of these new media forms. This is hardly accidental; as new computing tools allow us to build more immersive simulations and more intricate digital models of characters and their movements, digital humanists are adapting these tools and frameworks to produce serious games that have the capability of both altering the behavior of gameplay and studying that behavior in granular detail. This is the impact of applying computational thinking to research; as Jeannette Wing has argued, the computer as a human-built machine has encouraged human beings to attend to their own abilities to work heuristically toward solutions in the face of uncertainty. For a variety of reasons that are more germane to cognitive psychologists than humanities researchers, human beings are much more likely to think creatively, imaginatively, and analytically when confronted with rule-governed game play than they are when confronted by traditional assessment media that more closely resemble tests. Moving our work – both research and teaching – into the more creative and more productive space of gameplay will enable many rich possibilities for the humanities as a discipline.

Other NEH funded projects have traversed the arena of game-based research. In recent years, other projects have sought funding for gaming by building simulations of historic events such as battles from the Civil War, by constructing immersive worlds to reproduce distant environments such as Delphi and Bangkok, and by using hand-held gaming platforms to allow users to navigate physical spaces as though they were virtual. Such projects emphasize the vast potential for researchers working at the intersection of gaming and the humanities. Unfortunately, these projects become essentially single-purpose acts. Each initiative exhausts its resources in the completion of its model, leaving the funding agencies with no way to scale the enormous investment in infrastructure. Each project, while worthy in its own right, fails to move gaming into the mainstream of humanities research. We want to address that deficiency.

Our project aims to create a sustainable infrastructure for routine applications of serious gaming in humanities research and teaching. We propose to unite interested but unspecialized humanities researchers with game developers, new media faculty specialized in the study and production of gaming media, computer scientists studying the environment of games as a decision platform, and digital humanists committed to research at the juncture of computational methods and human culture. In the three-week institute, we'll undertake together an intensive investigation of emerging models for the application of serious games to academic work. The institute will push its

participants to question how gameplay can transform the intellectual spaces of academic work, and will draw from its participants' various interests and specialties to develop overlapping areas of concern and support. Since it provides both the space for a sustained discussion of the theoretical dimension of gaming and a supportive environment for the recursive hands-on work of building games, the project will ask its participants to grow and develop along two directions simultaneously. We'll also provide a year's worth of infrastructural support and technical assistance for the participants as they move back to their home institutions. We recognize that the institute can thereby provide the occasion for both intense engagement and sustained thinking that is then transformed by the participants as they integrate their experiences into the work of their peers at their home departments and programs. We want to assist that transition, by continuing to connect the participants, by hosting their projects going forward, and by supporting their efforts to translate their experiences for their colleagues and peers. By allocating a significant portion of its resources for technical support and infrastructural assistance during the year after the institute meets, our project can build discrete, self-sufficient yet inter-connected nests of faculty researchers working in humanities gaming. In this way our proposal represents a prudent use of the agency's limited resources.

### **Institutional Profile**

The University of South Carolina is fortunate to be able to offer the wide variety of faculty and staff needed to undertake a project of this size and complexity. With an active and ambitious Center for Digital Humanities, a Computer Science program eager to interrogate problems in the humanities, and a Media Arts department that values both producing media art and understanding its meaning and value, our institution can supply the specialists needed to undertake a project of this scope and size. During the three-week session, we will take advantage of the facilities of the school of Computer Science, which has purpose-built classrooms designed to facilitate collaborative work such as ours. Within each of the sponsoring departments, we are well prepared with the computing facilities to support twenty participants during the three-week institute; we intend to take advantage of this redundancy to provide multiple overlapping venues for small group interaction, after hours sessions, and weekend development. In the subsequent year, the Center for Digital Humanities will provide the infrastructure to host the participants' games and projects, through a shared server, allowing the participants to extend the work of the institute far beyond its temporal and spatial confines. We will happily support our core of twenty participants, and just as gladly extend the same support to new partners as they emerge. Facilities are not the limiting feature at our institution to an endeavor such as this; we look forward to sharing in the rich intellectual milieu afforded by the institute. The university is able to make available conference-style housing for the participants in the student dormitories, and there are numerous hotels within a few blocks of the university.

## Curriculum and Work Plan

The Institute is framed to give participants ample opportunities to work at the forefront of both critical theories of gaming and emerging toolkits and platforms for game building. We'll structure these parallel movements to overlap and work in tandem to sustain a rich level of inquiry for the participants. Instead of attending first to game theory and then to game development, our group will work to bridge the concerns of the two realms. For example, on a day when participants investigate the rule-governed nature of gameplay, we will also work with a highly structured rule-driven serious game built by Linguistics researchers to investigate questions of syntactic meaning. By moving between the two mutually supporting areas – development and theory, form and content – we hope to be able to employ the methodologies of each to interrogate the concerns of the other. In an ordinary six and one-half hour day, we imagine we'll spend roughly three hours on each of the two activities, with local differences dictated by the questions and concerns of the group. Many of the days have two such transitions, allowing multiple encounters within each area of work every day.

Our model divides the Institute into three weeks of work (for a detailed overview, see Appendix I), each with a separate theme for the game development, game play, and theoretical readings and discussion. The first week, under the rubric of the theme of "ReCreation: Making and Playing," will work to help participants quickly build familiarity with the five platforms and toolkits for game development: Portable Gaming Devices (PSP and iPhone), web-delivered and local Flash-based Games, the opensource Games4Change toolkit, and the two 3D game building platforms, Panda3D and Torque Game Builder3D. Simon Tarr and Duncan Buell will introduce the participants to the platforms and toolkits, providing ample opportunity to investigate gameplay under each platform. At the same time, participants will begin to investigate the cognitive and structural elements of game design, reading essays by some of the leading experts in game design. In addition to our interdisciplinary panel of project directors, participants will be able to learn alongside one of the foremost national experts in game design, Tracy Fullerton, author of the widely adopted *Game Design Workshop*. Fullerton, who literally wrote the book on the academic study of game design, will lecture on the structural elements of game design, such as rule governance and rule breaking. In her four lectures, Fullerton will also introduce central themes in the cultural theory of game design, investigating with participants the impact of gender roles on game play and the mechanisms of community building within gaming. Each day, participants will discuss foundational texts in the cultural operation of gaming, play exemplary serious games developed on our selected platforms, and work hands-on with the modular tools of game building to develop a fluency with the challenges and dependencies of the practice. In addition to the lectures by Fullerton, during the first week the co-directors will work to build a free-ranging discussion that begins at the elements of game design and works outward to engage the theoretical dimensions of transgression, recursive self-development, visual communication, and collaboration.



In addition to the lectures and discussions of game theory, the first week will structure an appropriate amount of game play to help participants make informed decisions about the subtle distinctions between the platforms and toolkits. The goal of the Institute is to assist the participants in building serious games for humanities teaching and research, so it is crucial that they quickly become familiar with the toolkits and platforms for building a serious game. The co-directors have collaborated to build exemplary games for the participants to interact with in the first week, games which can easily be mastered and whose development is illustrative of the potential for serious gaming in academic spaces. In addition to our easily-built sample games, we've arranged for faculty throughout the humanities to attend the institute and present research games and learning games to illustrate the breadth and depth of gaming as both a digital and traditional area of humanities scholarship. In the first week, we'll have faculty from Linguistics, Literature, Information Science, and Anthropology present academic implementations of serious games to the group, seeding the institute with cross-disciplinary activity that can help sustain transformative work.

At the end of the first week of the institute, our participants will have read and discussed some twenty essays and excerpts on game theory, game design, game play, and cultural theory. They'll have experience with at least ten different games, built from five different toolkits and played on several different platforms. Our participants will spend approximately 10 hours of hands-on work building proficiency across the toolkits, and will receive ample time to work one-on-one with a skilled staff member in order to develop a specific implementation plan for building a serious game. The first week of the session is critical, building a vocabulary and skill set as a foundation for the entire endeavor. For this important week, we've assembled a diverse team drawn from our multiple strengths and resources to ensure that participants are well-prepared to succeed in their work.

In the second week of work at the Institute, we'll transition from maneuvering through the generalities of game development as a set of structural problems and begin the process of building discrete games with our participants. In order to facilitate the many discrete processes of game building, we'll attend closely to games *as play*, interrogating the interactive nature of game play as a series of processes with multiple meanings and experiences. Our theme for the week's work, "Power and Structure," underscores the ways in which the processes of gameplay can coalesce into structural elements which have significant affective force for the players. To help our group focus on the processes of gameplay, we've selected an array of readings to discuss that include Michel Foucault, Manuel Castells, Alex Galloway, and Ian Bogost. In addition to our interdisciplinary team, this week will feature several guest lectures by Ian Bogost, a nationally-recognized rhetorician and game theorist and author of *Persuasive Games*. Perhaps best known for his innovative application of procedural rhetoric to the gamespace, Bogost's analytic methodology and theoretical underpinnings will help stimulate a fruitful questioning of many of the often taken-for-granted aspects of gameplay. Bogost's lectures will serve as a valuable component of the intellectual work

of the institute, helping us to understand gameplay not as an event but as an evolving set of practices and concerns, each of which is alterable.

Alongside the nuanced theoretical examination of gameplay afforded by Bogost and the other authors we'll read in the second week, our participants will work in a detailed and patient way through the elements of game building. No longer developing in unison, our participants will split into teams to work closely with our developers to implement the interactive processes of gameplay into their evolving games. Each team will be structured to allow them to build proficiency in one toolkit or platform. Some of our teams will work to construct a multi-user game environment using one of our opensource toolkits, structuring the  $n$ -dimensionality of action into the game design. Others will experiment with creating 2D and 3D textures and landscapes as expressive elements of gameplay, background components which work suggestively during the game to structure and constrain a player's decisions and affective. In tandem with our staff and team of developers, participants will learn to build modular structural elements of gameplay – for example, avatar creation, which entices users to map their own role-playing capabilities onto the actions of game characters. We'll attend closely to questions of player interface design during this week, often the most recursive element in game design. Throughout the week, we'll continue to provide several hours of hands-on activities, including one-on-one work with developers to produce the participants' serious games. Throughout the entire session, we'll provide lightly-staffed optional time in the labs to continue development on the weekend and in the evening, facilitating participants' disparate work routines.

As the session develops, we'll continue to structure play into the work of the Institute. As the participants' games evolve, the needs and abilities of our group will rapidly push beyond the limits posed by our sample games; instead of working from the samples, our group will start to model game development work on more established serious games that are freely distributed (see, e.g., [seriousgames.org](http://seriousgames.org)). Bringing in these sources as powerful examples of the possibilities of gaming, we'll draw on examples such as the Eternal Egypt application that provides not just access to the cultural artifacts from Luxor but also a virtual environment that simulates cultural experience as gameplay. Throughout the second week, each day we'll interact with a well-known serious game in order to provide a diverse background for our own game development, discussing the game's interactive elements and its ability to structure and facilitate our responses. We'll also continue bringing in faculty research and learning games from around campus, with projects from Cultural Anthropology, Ecology, Composition, and Philosophy. By sampling from well-developed and extremely successful serious games, we want to continue to push our group to imaginatively respond to the possibilities that games offer.

Our third week, "Form and Content," works to complete the games as digital projects and build a sufficient engagement with the work of serious gaming to sustain the participants as game developers at their home institution. We're fortunate to have

noted game developer and game theorist Ben Sawyer as a guest lecturer for four lectures during the third week. Sawyer is best known as the Co-Director of the Serious Games Initiative sponsored by the Woodrow Wilson Center for International Scholars. Sawyer's work blends game design, game theory, and learning or training games. Sawyer will lecture on his development of a taxonomy for games that connects gaming practices in corporate environments and e-learning platforms to pleasure-based casual gaming. Arguing for a continuum in gaming practices between formal and informal spaces, Sawyer will challenge the group to include as gaming practices that seem removed from some of the arena of play. Alongside Sawyer's lectures, the co-directors will work to examine the complex terrain of work/play spaces that participate equally in social milieu formation. We'll look closely at how games facilitate the conditions for human self-understanding within gameplay and at game development in and as education. In participating in the discussions that provide the interpretive framework for the third week, the team of co-directors will work to foster a series of synthetic moments that bring a sense of accomplishment to the session, while also enabling inquiries that move beyond the confines of the Institute to continue at the participants' home institution.

Outside the discussions that frame the game development, the group will focus on work that will deliver a completed game for the participants of the group. Participants will spend more than half of each session in intensive development workshops, working closely with the staff developers to complete the game projects in a sustainable way. We'll include learning sessions on the hardware options that the Center for Digital Humanities provides, including multi-player game hosting, Flash-based game delivery, and ongoing technical and design assistance. We'll work to ensure that the participants constitute themselves as a community, returning to the larger group to share their experiences and connect to their peers. We'll end the session by showcasing the successes, analyzing the deficiencies, and thinking through the processes together.

We envision the Institute running on a three week schedule, from May 10 to May 28, 2010.

## **Participants**

We will conduct a vigorous nation-wide recruiting campaign for faculty members and senior graduate students to attend this Institute. We'll recruit through the ordinary Digital Humanities points of contact—HASTAC and centerNet, but we'll also reach out specifically to smaller schools that have no digital humanities infrastructure. We'll ask applicants to include in their letter of application a description of the research or teaching application for which they wish to develop a game. We'll also ask them to describe the infrastructural support for Digital Humanities research at their institution. We'll prioritize applicants from schools that lack facilities to support game development and infrastructure. All applications will be blind-reviewed by a committee composed of the co-directors of the project.

## **Impact and Evaluation**

This project aims to take an emerging practice within digital humanities research and teaching – the use of games to facilitate research and learning – and render it supportable at institutions that lack a significant infrastructure to conduct digital humanities work. If successful, its impact will be measured by the number of institutions that can undertake emerging work in the digital humanities. The real work of the project, then, occurs not just during the three-week institute, but also during the year following the institute. We’ve devoted a significant portion of our budget – almost all of our salary expenditures occur after the institute has ended. We’ll work carefully to connect with our participants and support their work going forward, to enlarge some of the more sustainable practices within digital humanities research.

## **Staff, Faculty, and Consultants**

Duncan Buell, Principal Investigator, is the Chair of the Department of Computer Science Engineering at the University of South Carolina. His research interests include pure mathematics, pattern recognition within data, and facilitating undergraduate research. Dr. Buell is quite experienced in building and delivering serious games for education and training. In the past, he has worked with faculty working in Law and Business to develop game models for learning. For the Humanities Gaming Institute, Dr. Buell will lead the game development team.

Randall Cream, Co-Principal Investigator, is Associate Director of the Center for Digital Humanities at the University of South Carolina. His research interests include educational use of games, data mining and behavior modeling, and critical pedagogy. Dr. Cream has experience in serious games research, using gameplay decisions to model cognitive impact of the game. In the past, he has worked with students from across the university building learning games and studying causal games as learning moments. For the Humanities Gaming Institute, Dr. Cream will work with both the game development team and the theory and design team.

Heidi Rae Cooley, Co-Principal Investigator, is a theory-oriented new media scholar in the Department of Art/Media Arts and Film and Media Studies Program (University of South Carolina). Interested in the articulation of poiesis (creative production), aesthesis (sensory knowing), and ethos (practice of living), she will invite Institute participants to conceptualize serious gaming in terms of an ethical engagement, one that is attuned to the inter-relations among technology, sociality, and living bodies. Dr. Cooley will facilitate several of the reading discussions, participate in the discussions of game design, and enjoy the game play. Drawing on a notion of “eco-logical praxis” (Felix Guattari), she will suggest that gaming interpreted and pursued along these lines evolves in a dynamic, more intuitive manner because it recognizes and attends to our condition of always being in-relation. Points for discussion drawn from this

philosophical perspective will strive to imagine serious games as the condition of possibility for destabilizing the politicization of life attributed to modern governance of populations.

Simon Tarr, Co-Principal Investigator, is a media practitioner and theorist in the Art Department's Media Arts and Studio Arts division at the University of South Carolina. Prof. Tarr will be providing crucial design support for the Institute participants as they work toward developing their own serious games. Well-versed in several of the gaming platforms to be introduced, as well as familiar with the aesthetics of animation and gaming, he will challenge participants to think more complexly about the interface of their games. Moreover, given his theorizing of interactivity, he will invite participants to imagine serious gaming as a mode of performative enactment of ideas.

Ian Bogost, Guest Lecturer, is a videogame researcher, critic, and designer in the Literature, Communication and Culture at the Georgia Institute of Technology. Dr. Bogost will introduce game criticism that contextualizes games in the long history of human expression. An acclaimed game rhetorician, he will discuss how games make arguments. Interested in uses of videogames outside entertainment, including politics, advertising, learning, and art, he will encourage participants to explore the richly diverse possibilities onto which gaming might open. His analytic methodology and theoretical underpinnings will help stimulate a fruitful questioning of many of the often taken-for-granted aspects of gameplay. Bogost's talks will serve as a valuable component of the intellectual work of the institute, helping the participants to understand gameplay not as an event but as an evolving set of practices and concerns, each of which is alterable.

Tracy Fullerton, Guest Lecturer, is a nationally recognized gaming expert and designer in the Interactive Media Division of the USC School of Cinematics Arts. Fullerton will present on the formal elements of game design and introduce participants to the emerging critical discourse around the nature of meaningful gameplay. In a series of talks, she will provide a necessary overview of the central themes in the cultural theory of game design, investigating with participants the impact of gender roles on game play and the mechanisms of community building within gaming. A proponent for a playcentric approach to design, she will challenge participants to think through gaming and to conceive of gaming as theoretical praxis.

Ben Sawyer, Guest Lecturer, is a nationally recognized expert in serious gaming. Co-Director of the Serious Games Initiative at the Woodrow Wilson Center for International Scholars, Sawyer is an expert in game design, learning through games, and socially conscious gaming practices. An early and active member of the serious gaming community, Sawyer will be instrumental in providing leadership to participants who wish to deliver small, portable, engaging games that tackle relevant social issues. He will lead four guest lectures and work with participants on game development.

Jun Zhou is Lead Developer at the Center for Digital Humanities at the University of South Carolina. She will work closely in all aspects of game development. Most importantly, she will direct the support efforts for the extended support period of the project. Zhou is an experienced game developer and talented programmer.

### **Budget Notes**

There are two noteworthy elements of this institute's budget. First of all, it spends an extraordinary amount of money on graduate and undergraduate student support. These expenses are justifiable because this project, unlike most of its competitors, aims to greatly extend the capacity to participate in research in the digital humanities. With that aim, participant support becomes not an adjunct to the project, but rather the core of its work. Well-trained students are ideal for providing the support necessary, since they will benefit from having to learn to communicate with non-specialized audiences. These same students will be present at the institute, allowing a connection to develop between participants and the students who will help support and sustain their research.

The other extraordinary expense is the stipends for participants, offered at the standard NEH rate. We're hoping to attract researchers from institutions which offer less infrastructural support to their faculty than do many others. We believe that we can essentially subsidize sustainability by sharing knowledge resources. While it would be fiscally irresponsible to develop multiple competing centers with duplicate infrastructures, much of the research in digital humanities relies on large-buy-in infrastructure. Just as we want to extend participation to new partners by sharing knowledge and computing infrastructure, so too do we want to encourage participation by subsidizing subsistence costs. We estimate that most applicants will fully spend their stipends on subsistence costs to attend the institute.

## The Sapheos Project: Transparency in Multi-image Collation, Analysis, and Representation

**Competition: NEH Start-up Grants in the Digital Humanities**

**Submitted: April 8, 2009**

**Awarded: \$50,000 August 2, 2009**

**sapheos.org**

Randall Cream, Project Director

Song Wang, Co-Director

The Sapheos project emerges out of my work as project director for the Spenser Archive, where I encountered the problem of collation firsthand. Sapheos, a Greek word for 'transparency' or 'clarity' (with rhetorical overtones), is a suite of middleware applications, most notably a digital collator. My solution to the problem of collation is to use transparency to overlay images of multiple texts over each other, forming an assembled text from the multiple. Through transparency, uniformity emerges as a single, clear text; variants emerge as visibly unclear spots that are easily identified by the software. This is hardly a novel approach-- most collators work by visually aligning or superimposing images. Until now, digital collation has been elusive: images are simply too "messy" to align computationally. Sapheos achieves this alignment not just for a pair of images, but for larger multiples (alpha testing worked with six simultaneous copies).

Our startup grant merges my solution with the work of the Computer Vision team at South Carolina, led by Song Wang. Wang's group uses MATLAB to apply algorithmic solutions to difficult problems in computer imaging. Our prototype is already functioning in the code; we are working on automating a process that is still manual. At present, we are approaching 90% reliability.

The Start-up grant is a good fit for this project because the competition is meant to fund prototypes for future grant categories. We envision pitching Sapheos to several additional competitions this year.

## The Sapheos Project: Transparency in Multi-image Collation, Analysis, and Representation

For many in the humanities, the book is irreplaceable as both a work of art and a material aspect of human culture. Already a hopelessly outdated technology when Augustine wrote his *Confessions* in the 4<sup>th</sup> century—tied to the metaphor of the scroll and purposely built to mimic the mechanical act of scrolling from left to right, top to bottom—the codex has survived as bearer of accident, meaning, and metaphor. More recently, digital humanists have worked to incorporate the metaphor of the book into digital devices for reading, from the British Library’s Silverlight-based page turning application to Kindle’s digital ink. Reading as thumbing through a codex, scrolling through page-images sequentially, remains a key interface requirement for many users, writing the metaphor into a new array of technologies.

Central to the book, then, is the ordering of a collection of pages, each separate and distinct, yet linked by the logic of the fold or the gathering. The digital humanities remains relatively under-developed in responding to this aspect of reading culture. Much of our recent focus has been concerned with developing efficient and accurate methods of text markup—while TEI is now the standard means of storing and representing textual data, the most widely used schemas for markup eschew the page as an ontological entity within the xml. The marked-up textual data contained within books dominates current methodologies, reflected by metadata, text mining and aggregation, and textual analytics, disregarding the page as anything more than a mute bearer of meaning. However, many early-modern cultural artifacts, such as Shakespeare’s Quartos and Folios, Spenser’s 1590 *Faerie Queene* and *Shepheardes Calendar*, and the vast majority of medieval manuscripts, cannot be represented adequately by marking up their textual data within XML tags—features such as handwriting style, textual variants, textual decoration and embellishment, and accompanying figures remain only barely conformable to the TEI standard, and the visual representation of this information is not assimilable to a word-based markup. In many cases, nothing less than the ability to interact with images of the pages will suffice.

The Center for Digital Humanities at the University of South Carolina, partnering with the Computer Vision Group in Computer Science Engineering, is developing a suite of applications that responds to this need for software that reunites text and image of text in an intuitive manner. Our Sapheos project ([sapehos.org](http://sapehos.org)), led by Randall Cream, associate director of the CDH @ SC, unites two of these applications into a powerful apparatus for interacting with early modern books. Our most ambitious project is our digital collation software, prototyped in MATLAB and delivered as an open-source project using C code. This software builds on existing projects and methodologies to deliver both a back-end collation tool and a powerful front-end interface for interacting with large datasets of books, a common occurrence for the best-known authors of the early modern period. An equally significant component of our Sapheos project is our software for automatically sectioning and generating (x,y) coordinate pairs for page images. Written in Java and being ported to MATLAB for C compilation, this software is designed to take images of pages with existing XML markup and insert (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.

These two applications are linked in their ability to allow the Sapheos project to deliver image and text as coequal components of early modern books. 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. Together with the collation software (see screenshots



pp. 25-29), these two tools have the potential to transform scholarship on medieval and early modern print culture.

### Enhancing the Humanities through Innovation

The Sapheos project contributes greatly to innovation in the humanities in three key ways: (a) in user interface, by producing a means by which users are able to interact with displays of large image-based datasets; (b) in optical collation, which remains a time-consuming and costly affair for scholars, as well as destructive for fragile primary materials; and (c) in bringing the flexibility of xml-based data into image-centered projects. The first two innovations concern the digital collation software; the third innovation is produced by the (x,y) coordinate software.

Our collation method draws from existing best practices in the field (see Environmental Scan, page 5) in order to produce meaningful results that can be used by scholars familiar with collation in early modern textual studies. 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*, 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.

In “The Notion of Variant and the Zen of Collation,” from *The Myth of Print Culture*, Joseph Dane interrogates the collational impulse as only partly arising out of any uncertainty within the text itself. Instead, Dane asserts that the process of collating witnesses (copies) exists as much as a feature of the bibliographic impulse, as a sort of “archive fever” for the textual scholar. Randall McLeod, long an innovator in collation technology, agrees that intra-edition variation exists as something of a mixed bag between meaningful difference, accidental error, and undecidability. Editorial practice, from McLeod’s point of view, ought to fully represent, rather than disentangle, such undecidable moments.

With intra-edition collation as an inextricable element of humanities scholarship, the innovation of the Sapheos project’s computer-assisted collation is greatly clarified. While the siren call of optical collation has resulted in several other projects building competing methodologies (see Environmental

Scan, below), our methodology innovates in several key ways. An outline of our methods will demonstrate this innovation.

Like all optical collation, Sapheos works by aligning images to suggest variance. Whereas some manual collators use the mind's ability to stereoscope two visible fields, most computing methods superimpose page images to produce the same visible effect. Using a linear opacity algorithm, images of the same page are combined to form one visible page, with variants easily detectable as "fuzzy" or out of focus areas within the page (see figure 2, Appendix I, page 26). Our innovation to existing approaches is two-fold. First of all, we extend collation beyond the binary, scaling to four and five simultaneous copies in our alpha version and extensible to as many as eight copies in manual proof-of-concept testing. (see figure 1, Appendix I, page 25) Secondly, our collation method works with off-the-shelf images of witnesses. This is a significant and important innovation, since most computer-based methods encounter insurmountable obstacles when comparing images taken at different resolutions, with different optical characteristics caused by lens distortion and focal field difference, and with the infinitely variable deformity of page curvature. Instead of manually producing uniform images by manipulating early modern books and thereby risking damage to priceless artifacts, Sapheos software deforms existing images, in all of their peculiarities. Our partner in this strategy is the Computer Vision team at the University of South Carolina. Accustomed to working with bio-medical imaging applications, our collaborative approach to the problem of images of multiple copies with idiosyncratic properties by applying methodologies derived from working with images of the human body in motion. Rather than approaching difference as binary—where image x is different from image y in discrete ways—we treat difference as a feature of *each* image, deforming and aligning each image to some intermediate position generated from the set (see Work Plan, page 7). The result is the ability to "stack" or layer multiple copies, each aligned through progressive deformation to a new state. Difference is instantly and visibly apparent as an easily noticeable blur on an otherwise crystal-clear page, represented with remarkable clarity, for any image within the stack.

By bringing multiple copies into the process of collation, Sapheos produces not just a back-end for collation but also suggests a powerful user interface for interacting with archival projects with multiple copies of extremely similar materials. Still in the planning stage, our user interface will allow individuals to select witnesses to layer together into sets, combining the images to produce an assemblage that approximates unity. As a user interface, Sapheos brings dissimilarity into focus by rendering agreement mute—similar texts, when layered, merely appear as *one* text, whereas differences between the layers manifests as visible blurs, disrupting the text's singularity (see mockup, Figure 2 Appendix I page 26). As a collation backend, Sapheos is a powerful tool for textual bibliographers; as a user interface, Sapheos has the potential to allow users to meaningfully interact with large datasets in ways that highlight and refine the characteristics of the collection. As part of the user interface for interacting with large datasets, Sapheos software links the image stacking algorithms to a sophisticated page-turning applet that loads underlying xml from the pages represented, thereby serving to facilitate a variety of means of interacting with books held within digital archives.

## **Environmental Scan**

The ubiquity of the desktop computer brought a wide-eyed optimism to humanists in the early 1980s, when the promise of automating the difficult task of collating appeared to be just around the corner. On the Rare Books and Special Collections forum on Bitnet, we can easily find archived discussions

that attest to the dream of digital collation software. Of course, these expectations proved to be unfounded. Early experiments in using a computer to collate proved more or less as unwieldy as mechanical collation, with the added difficulty of accurately processing the images needed to digitize the copies to be collated.

Currently, collation software is very well developed for text-based projects, with an array of tools such as NINE's *Juxta* and Susan Schreibman's *Versioning Machine*. These flexible applications allow users to collate differences between well-marked up texts, essentially mining the two copies for textual difference. For early modern texts, though, such software is no solution to the problem of interacting with intra-edition variation. With OCR being notoriously unreliable with early modern type layout, optical collation remains the only available method.

The utility of optical collation is well-established by the few other projects that have taken up a similar problem recently. While there are an array of image comparators from commercial or open source projects—such as Bolide Software's Image Comparer, Tigris's TortoiseDiff, Adobe's Photoshop, ImageMagick, and Virtual Lightbox by Matthew Krischenbaum and Amit Kumar—none of these software solutions is adequate to correct for the multitude of differences that are a result of two copies from an edition having very different individual histories and provenance over a 500 year span. Simply put, available images of witnesses from a common edition reveal as much difference as similarities, highlighting the multiple attempts at rebinding and trimming, the vagaries of fading and coloration, and the almost insoluble problem of page curvature from tightened book spines.

Despite the proliferation of text-based humanities work, there are two noteworthy projects which take up the issue of digital collation. The most ambitious and well known of these two projects is the HUMI (Fumi) project launched almost a decade ago by Keio University, Tokyo, Japan. Working with the British Library, the Pierpont Morgan, and other prominent partners worldwide to digitize and collate the known copies of the Gutenberg Bible, faculty of the HUMI project, led by Toshiyuki Takamiya, developed innovative methods for working with both the single-volume and double-volume instances of this famous text. As detailed in an article in the Japanese-language *Journal of Library and Information Science* 53.3 (2005), "Toward Collation with Digital Images," the HUMI project developed software to collate these 15<sup>th</sup> century texts by superimposing the images and aligning the pages to produce difference as a visible blur, in much the same way that Sapheos software layers images. There are a few significant differences, however. As detailed by project member Mari Agata, HUMI software relies on manually flattening the images of the bibles using bamboo sticks to generate conformable images. Despite the care taken by project members, we believe this approach is unduly risky to valuable holdings and relies too greatly on care and expertise. Our approach to the same problem is to automate deformation of image sets algorithmically, thereby allowing the underlying artifact to remain undisturbed by human intervention.

Another promising candidate for digital collation is software being developed by a team of researchers at McGill, Harvard, and Geneva Universities, *Aruspix*. *Aruspix* is purpose-built to work with early modern music, collating manuscript and machine-printed music texts. One of the innovations of *Aruspix* is to represent divergent witnesses by using different colors to differentiate the content of each copy. Like Sapheos and HUMI, *Aruspix* works to perform digital collation by overlaying images of texts to make difference visible for users. However, due to project idiosyncrasies, *Aruspix* performs OCR as a crucial step prior to collation. This OCR step allows the project to collate between editions,

not just within editions, but it fundamentally constrains the applicability of the project's software—musical notes are relatively easy to differentiate, whereas early modern printed books routinely use ligatures, non-standard spacing, and embellished and decorative characters. An OCR-based collation can never isolate intra-edition variants as effectively as optical collation, and so remains an unsuitable application for textual scholarship.

Sapheos software addresses these deficiencies and adds a substantial capability that enhances its innovation: collation of multiple copies. By extending the process of collation to between four and eight copies simultaneously (see Figure 1 Appendix I page 25), Sapheos brings remarkable efficiency to existing models of digital collation and representation.

### **History and Duration of Project**

Closely aligned with the Spenser Project ([spenserarchive.org](http://spenserarchive.org)), with partnering faculty at institutions such as Cambridge University, Washington University St. Louis, Pennsylvania State University, and the University of Virginia, the Sapheos project benefits from a strong association with established scholars working in digital projects in the humanities. As an early-modern editorial project, the Spenser Project has long recognized the need to conduct collation of early modern texts. Beginning in 2005, project members began developing digicoll, an application written in C++ to identify textual differences between image pairs. Digicoll, led by Craig Thomas and Aaron Zeide at Washington University St. Louis, used an open source OCR library, claraocr, to perform glyph recognition and page comparison. For the reasons outlined above, an OCR solution remains unsuitable, leading to software redevelopment. Led by Randall Cream, a new solution was devised in the Fall of 2007 and project development resumed with a robust collaboration with the Computer Vision Team at South Carolina.

With an initial proof-of-concept demonstrated in Photoshop, development on the algorithms necessary to collate n-n copies began in MATLAB. Our initial alpha-type software, demonstrated in Appendix I (Figure 3 page 27), achieves its results by manually selecting the registration points for image deformation. Work on the underlying mathematic models for automating the detection and processing of registration points for a number of images continues, subject to the vagaries of funding models.

Our (x,y) coordinate software remains a much more developed software package than the collation software, yet for continued development it must be ported over to MATLAB in order to integrate and deliver the application as an open source project written in C.

Looking ahead, we plan to continue development of Sapheos project software, subject to funding models, over the next several years. With a publicly delivered beta solution, we'll move aggressively to build a user interface that takes advantage of the flexibility of the underlying software. We anticipate the total time to a public beta to be approximately 15 months, with a stable, feature-rich application to be completed in the range of 24 to 36 months.

### **Work Plan**

Sapheos software is jointly developed by the Center for Digital Humanities at South Carolina, led by Associate Director Randall Cream, and the Computer Vision Team, led by Song Wang. Our

development of the software is anticipated to occur over the next 15 months. Working as a team, the Computer Vision Team serves as a resource for the Sapheos project, with a collaborative work structure and many opportunities for problem solving and joint critique. The project director, Dr. Randall Cream of the CDH @ SC, will use a course release each semester to directly supervise the project and ensure work is completed in a timely fashion. In order to participate fully in the Computer Vision Team, Sapheos will sponsor a graduate student developer. We will also sponsor two undergraduate students to work on the project, developing the user interface, software testing, and documentation. Jun Zhou, Lead Developer at the CDH @ SC, will coordinate between the undergraduate and graduate developers, working closely with the Project Director to ensure that the development schedule is maintained. Our work plan to develop this tool includes

1. Software design and implementation: 34 weeks
  - Image preprocessing: 2 weeks
  - Separating printed contents and handwritten contents: 10 weeks
  - Extending finding x-y coordinates of text lines and words: 6 weeks
  - Image overlaying: 10 weeks
  - Difference detection: 2 weeks
  - Data visualization: 4 weeks
2. Software testing and revising: 15 weeks
3. Document and packaging: 5 weeks

### **Project Staff**

The Sapheos Project is fortunate to benefit from a variety of substantial relationships. Core project staff for software development include Dr. Randall Cream, English department and Center for Digital Humanities; Dr. Song Wang, Computer Science and Computer Vision Team leader; Jun Zhou, Lead Programmer at the CDH @ SC; Jarrell Waggoner, Ph.D. student in Computer Science and key personnel, Computer Vision Team; Jon Bolt, undergraduate Computer Science Engineering major; Ekshita Kumar, undergraduate Computer Science major; and Dr. David Miller, PI, Spenser Project. Cream, Wang, Zhou, and Waggoner will meet on a weekly basis for the duration of the project, ensuring that the team deploys its various strengths in solving unanticipated hurdles. Each of these team members will work for 10 hours a week over the course of the academic year. The undergraduate students, Bolt and Kumar, will devote on average 5 hours a week on the project over the course of the school year. As a textual scholar and a renaissance humanist, Miller will serve as a valuable resource or discussing and identifying difficulties and solutions in collating and representing early modern texts.

### **Final Product and Dissemination**

Our project will release all of its work as opensource code, encouraging other researchers to use our work, benefit from our investment of resources, and alter our code to extend the benefits of our research. We'll present our work at DH2010, Computer Vision 2010, and TEI 2010. We'll release a whitepaper at the end of the grant detailing the lessons learned in adapting Sapheos software to current problems in humanities computing.

## Human Voices Project

**Competition: Digging into Data, NEH/NSF/SSHRC**

**Submitted: LOI March 15, 2009; Final Proposal July 15, 2009**

**Proposed Award: \$300,000**

**Pending; decisions should be announced in December**

**[humanvoicesproject.org](http://humanvoicesproject.org)**

Randall Cream, Principal Investigator and NEH Project Director  
Michael Welge, UIUC, PI, NSF  
Paul Yachnin, PI SSHRC

The Human Voices Project is a data mining project that uses a feature of human behavior to introduce meaning into the inscrutable process of concept formation. Using inter-textual quotation and citation to create semantic networks from the articles in humanities databases—JSTOR, in this iteration—Human Voices allows existing scholarship to speak with a multitude of voices from the past, in colloquy with the present activity of the disciplines.

Human Voices Project is a multi-national, multi-institutional team of scholars. In addition to South Carolina, the team includes McGill and UQaM in Quebec, and UIUC and UCSB in the States. JSTOR, Google, and the NCSA are all contributing to our project. While the work will occur on numerous campuses, I'm proud to say that the grant narrative, innovative approach to data mining, and partnerships with UIUC/NCSA reflect my work. I'm proud (and humbled) to lead such an impressive team.

**Human Voices Project**  
**Randall Cream, University of South Carolina, USA, Project Director, NEH**  
**Michael Welge, SEASR, NCSA/University of Illinois, USA, Project Director, NSF**  
**Paul Yachnin, McGill University, Canada, Project Director, SSHRC**

The Human Voices Project proposes a unique response to the proliferation of data in humanities computing. Instead of adopting computational analysis as a means of reducing complexity, Human Voices resituates the heterogeneity of the humanities by re-activating the interconnected nature of humanities work. Our project aims to mine data not for reduction but for analytic amplification by developing the inherent interconnectedness of humanities work. Data mining applications inevitably rely on reductive algorithms that analyze texts by flattening variables, identifying patterns, and reducing the ontological complexity of concepts in order to allow orderly computation to occur. Such models yield precise and useful results and enable pattern matching on enormous scales. Human Voices rejects these approaches, since humanities data cannot be made more meaningful by reducing its complexity. The recent Netflix Prize competition illustrates this difficulty: one million dollars and a team of several Nobel laureates only generated a 10.05% efficiency increase in predictive searching—suggesting that we are in an era of diminishing returns to algorithmic searching using keywords.

Human Voices re-imagines humanities computing with the humans in it. Our approach to the challenge of the proliferation of data is to not reduce that data, but to make it more human by attending to its inherent complexity. Our project works simultaneously with three archives in the humanities: (1) the English Broadside Ballads Archive, housing 120,000 records of ballads written between 1500 and 1700; (2) the Spenser Archive, housing the works of the 16<sup>th</sup>-century poet Edmund Spenser, and (3) JSTOR, the largest archive of humanities and social science scholarship in the world, housing approximately 5 million journal articles. Using automated citation extraction to reveal the multiple “aboutness” references that distinguish humanities scholarship, Human Voices uses data mining to multiply complexity by revealing affiliation relationships that activate semantic networks inherent in the data. Our work proceeds in three phases and draws on our partners’ complementary strengths: through automated citation extraction (Phase I), text segmentation into multiple overlapping units of semantic meaning using these citation units (Phase II), and semantic analysis through a process that flexibly attends to the multiplicity of layers and recursively weaves in ongoing usage of the data to refine meaning (Phase III).

Our team draws on the collaborative framework of a strong partnership with SEASR, a flexible platform for browser-based text and data analysis. Alongside the Making Publics project, a multi-institutional, multi-national team of researchers investigating public formation around works of art and intellect in the early modern period (1500-1700), Human Voices brings an innovative data mining component to this kind of humanities scholarship, allowing the underlying networks of affiliation, citation, and adaptation to emerge visually. The impact of our methodology—working with humanities material on a discrete, concept-level, thereby facilitating a massively self-proliferating network of affiliation relationships to consciously develop— is to enable a qualitatively different set of questions in the humanities to emerge. Rather than working *within* the text and then looking for connections to events *external* to those texts, Human Voices situates the primary and secondary materials of the humanities as participants in a series of overlapping conversations.

Our team is diverse, large, and multi-disciplinary, with private partnerships (participants from Google & Collexis) entering into partnerships with researchers from English, Computer Science, the Digital Humanities, History, Geography, and Art. On the US side, Randall Cream will serve as project director to a partnership on the NEH side of the project that includes a multidisciplinary team of researchers and a partnership with the private firm Collexis; the SC team will work closely with a team of researchers at the U of California at Santa Barbara, led by Project Director Patricia Fumerton. Also on the US side, Michael Welge will lead a team at the NCSA/UIUC developing SEASR. On the Canadian side, Paul Yachnin will serve as project director to a multidisciplinary team of researchers at McGill University, aided by a productive partnership with a Google software engineer.

## Project Participants

Auvil, Loretta, National Center for Supercomputing Applications (NCSA), University of Illinois-UC

Buell, Duncan, Computer Science Engineering, University of South Carolina

Cream, Randall, Center for Digital Humanities, University of South Carolina, **US, Project Director, NEH**

Eberhart, Marlene, Making Publics Project, Department of English, McGill University

Folkerth, Wes, Making Publics Project, Department of English, McGill University

Fumerton, Patricia, EBBA, University of California-Santa Barbara **EBBA PI and UCSB Project Director**

Leicht, Stephen, Chief Operating Officer, Collexis, Inc, South Carolina

McAbee, Kris, EBBA, University of California-Santa Barbara

Matthews, Manton, Computer Science Engineering, University of South Carolina

Meunier, Jean-Guy, Philosophy, Université du Québec à Montréal

Miller, David Lee, Department of English, Center for Digital Humanities, and Spenser Archive, University of South Carolina

Milner, Matthew, Making Publics Project, McGill University

Nebeker, Eric, EBBA, University of California-Santa Barbara

Petric, Vlad, Software Engineer, Google, New York.

Sieber, Renee, Geography and Computer Science, McGill University

Stahmer, Carl, EBBA, University of California-Santa Barbara

Welge, Michael, Director, Automated Learning Group, NCSA/University of Illinois Urbana Champaign, SEASR, **US, Project Director, NSF**

Yachnin, Paul, Making Publics Project, Department of English, McGill University, **Canada, Project Director, SSHRC**

Zhou, Jun, Center for Digital Humanities, University of South Carolina



## Human Voices Project

**Randall Cream, University of South Carolina, Project Director, NEH**  
**Patricia Fumerton, University of California, Santa Barbara, EBBA, Project Director**  
**Michael Welge, SEASR, NCSA/University of Illinois, Project Director, NSF**  
**Paul Yachnin, McGill University, Project Director, SSHRC**

The Human Voices project uses large-scale data mining to integrate humanities scholarship with primary texts in a flexible, re-configurable constellation of affiliation relationships. Using the common occurrence of inter-textual quotation and citation as a launching point to build dense nodes of affiliation, Human Voices resituates acts of scholarship into a multivoiced dialogue that allows multiple ambiguities of interpretation balanced by the intersubjective grounding of interpretable texts. Rather than subscribe to a notion of data mining as a practice that is inherently opposed to humanistic inquiry, Human Voices understands that the analytic methods of data mining already drive many of the familiar acts of humanities scholarship. The Human Voices project aims to displace serendipitous and accidental discovery as prime movers of humanistic interpretation. Our methodologies open up computer-assisted analytic interpretation to a broad array of scholars far removed from computationally intensive disciplines.

One key methodology within the Human Voices project segments the products of scholarship in the humanities by dividing the texts—articles in JSTOR, the largest archive of scholarship in the world—along the lines of citation and quotation. Our task—easy to describe, yet nevertheless a fairly dense computational endeavor—is to automate the identification, tagging, and segmentation of citations within the datasets. We aim to identify every sentence that cites an external work and insert the standard <cit /> tag to identify that citational act. This is a difficult enough task (see environmental scan, below) and we envision spending approximately half of the grant period refining our approach in ways that allow us to align our analysis with the vast contours of our dataset. It's easy to imagine the transformative potential of the results of such a machine-driven analysis. Articles that are now grouped by author, subject, and a few machine-extracted keywords, will instead be dynamically reconfigurable as ontological portfolios of a vast number of affiliation acts represented by citation and quotation. Instead of the paragraphs of an essay only constituting an irreducible whole, they will be able to enter into conversation with each other (through association) and with the paragraphs of the scholars and other texts (primary or secondary) that the scholars cite, as well as with the scholars and texts that cite them. Scholarship, thereby, becomes a human act of affiliation, and work in the humanities becomes more human centered. Or to put it more plainly, the voices of texts and scholars occasion, through our research paradigm, a conversation that speaks across time and space to engage in sustained deliberation and dialogue.

Our project, then, contains a highly-feasible yet transformative data analysis phase, a recursive modeling phase to adapt that analytic to conform to the contours of the datasets, and, importantly, a visualization and human interface construction phase where we develop an approach to data mining that employs the uniquely human qualities of judgment, analytic insight, and creativity through a recursive research environment. Human Voices uses the **Software Environment for the Advancement of Scholarly Research (SEASR)** in order to facilitate a recursive, human-centered approach to data mining that opens up data analysis to scholars who would never begin to look at code. Through a flexible platform that invites collaboration on analytic “flows” and visualization layers, SEASR allows the analytic might of data mining to operate through the browser window, opening almost any browser-readable data for analysis. Instead of asking users to take data to some external site for analysis, SEASR picks up data in the browser window and brings a fluidity to data mining by using a web-services layer that interacts with remote servers through a dedicated port. Open source, community-centered, and aggressively engaged with active researchers throughout the humanities and information sciences, SEASR is a significant platform for conducting computational analysis.

At its core, our project consists of three computational tasks of increasing ambition and difficulty. The most feasible element of our workplan is to automate the identification of citation references from our datasets. Although this is a non-trivial act (see environment scan, below, for some of the issues associated with citation identification and extraction), we're confident we can model this identification recursively throughout the JSTOR dataset during the 10 months of project time devoted to this process. The second phase of our workplan involves using a special subset of citations, quotations and paraphrases, to create units within the largest dataset, JSTOR. Each of these units will consist of the quotation and/or paraphrase, the citation, and associated sentences that explain and develop the referenced point. We'll construct the frameworks that allow these citational units to be cross-referenced with their cited texts and the texts that cite them, creating highly-related, densely associated nodes that represent the elements of scholarship that are most valuable, most controversial, or most noteworthy. Our longer term goal, outside the scope of this brief 15 month project, is to conduct semantic analysis on these units, identified with concepts, and allow concept mapping to span across and between essays as an assemblage of citation units.

This three-phased approach represents a responsible, extensible, innovative approach to the uniquely scaled datasets of interdisciplinary scholarship and research. It links definable, benchmarkable outcomes to high risk/reward methodologies, providing a justifiable return on investment for a project of this size and scale.

## **Datasets**

Our approach to data analysis in the humanities takes advantage of the enormous scale afforded by advances in storage density. JSTOR, the largest dataset of our four primary data models, is a five million article archive of scholarship in the humanities, sciences, and social sciences. Our three other datasets are much smaller by comparison, but together represent a significantly dense datapoint in humanities research. The English Broadside Ballad Archive (EBBA hereafter, <http://ebba.english.ucsb.edu>) is a culturally significant archive of approximately 5,000 English printed ballads from the early modern period (1500 to 1700), each multiply viewable as text and art in high resolution formats as well as in XML and MARC records, and with over 2,000 song recordings. MaPs brings together scholars based in Canada, the US, and the UK, as well as a network of over a hundred associates and correspondents, working in an array of human-centered disciplines, from Art History, Musicology, History, and Literary Studies, concerned with mining cultural archives and documents in order to trace the development of groups of association, markets, media and their impact on the formation of notions of publicity. The Spenser Archive is an archive of the works of the 16th century poet Edmund Spenser, with stunningly high resolution images of 16th and 17th century editions and highly structured markup of the texts. Together, the three early modern archives create a densely concentrated and diverse dataset for information about and from the early modern period. Our datasets weave together two very different forms of literate arts-- Spenser's epic poetry is at once quite similar to (in a formal sense) but also strikingly different from the ballads of the period, allowing the Human Voices project to interrogate very specific questions of aesthetics and semantic usage. By combining these two early modern archives with the MaPs project, which seeks to foster a collaborative exchange through social networking as a means of scholarship, Human Voices is able to develop a detailed and sustained inquiry into the poetics and politics of space in the early modern period.

Working with the primary archives, even with the additional layer of metadata from scholarly interaction provided by MaPs, is insufficient ground for the sorts of questions we want to facilitate for humanities researchers. Rather, using SEASR's browser-based platform, we want to perform analysis of the primary (historical) in tandem with the extremely large secondary dataset, JSTOR. The dense datapoints of early modern material should provide a sufficient node structure to build a useful affiliation network of

contemporary scholarship. JSTOR is key to this project because of its size and density. The five million articles in JSTOR represent one hundred and fifty years of scholarship, by no means a continuous and homogenous dataset. Unlike the early modern datasets, JSTOR contains texts from a variety of disciplines, from the sciences to the arts and humanities. Early scholarship bears little formal similarity to the models under current practice in the academy. The vast contents of the JSTOR archive represent a valuable opportunity to conduct model information retrieval within a dense yet varied dataset with somewhat limited or controlled vocabularies.

In structure, the JSTOR archive's data is dissimilar in important ways from the three other archives. The metadata of JSTOR is highly structured, well schematizing the bibliographic information about each article. The contents of the articles, though, are much less structured, generally marked up only by page tags to associate text with image. Processing the data of the JSTOR archive will more closely resemble working with unstructured text than mining the richly structured, highly tagged environment of the other archives. Our analysis tools, built from scratch using Python, Java, repurposed from existing code from NORA and the Monk workbench, and existing SEASR modules using D2K and UIMA, will readily adapt to both environments.

Working closely with the rich and diverse archives that constitute our dataset—EBBA, the Spenser Archive, and JSTOR—the MaPs and other project participants in this project can draw on their long history of looking at inter-textual and inter-scholarly discourse of the early modern period. Our robust collaborative relationship with JSTOR especially gives Human Voices unprecedented ease of access to the data for modeling and testing and at the same time a vested interest in developing tools that are portable to a more restricted model and therefore available for general usage. While project testing will use local copies of all of the data, finished modules will use API and a webservice model to make calls against remote data, allowing user authentication, session denomination, and other interface and security layers to be deployed. The Human Voices project recognizes the tensions between responsible access and the economics of preservation, and is bound by its agreement to respect the intellectual property contained within the archives of the dataset.

## **Environmental Scan**

The Human Voices project innovates not through the creation of never-before seen algorithms but by repurposing inherent capabilities from established areas of computing in new and exciting ways to produce transformative analytic results in the humanities. Our first phase, automated citation identification in academic scholarship, is a necessary pre-component to the analysis provided by our methods. Citation extraction is notoriously difficult to model due to the inherent ambiguities afforded by proper names of authors and journals. Just recently, in *D-Lib* 15.3 (2009), J.H. Canos et al propose adopting a unique identifier (DOI) in order to facilitate disambiguating citation references in scientific literature. Given the growth of extremely large datasets in scholarship (particularly in the sciences), ambiguity is a persistent and irreducible component of citation identification. Our chosen dataset of secondary scholarship, JSTOR, however, already employs a metadata field with a unique identifier tag, allowing precise references to build and inhere. We'll take advantage of this precision in our code, structuring relationships not between authors and texts but between unique identifier fields. The unique identifier tag of the JSTOR dataset give us unique, 1:1 relationships of citations and allows us to disambiguate extremely common journal names and surnames with ease.

Methodologically, Human Voices draws from several existing digital humanities projects that work in areas of citation extraction and identification. Perhaps the best known of these is the open source citation project, *OpCit*. An extremely ambitious project whose agenda far exceed the technical capabilities of its time, *OpCit* worked from the 1990s until 2002 to produce a standard for identifying, using, and referring to citations in the emerging world of text markup and text analysis. Their team, funded through the JISC

and the NSF and deeply influential, aimed to produce an open standard for archives, intellectual rights holders, and content aggregators to facilitate interoperable search and extraction. Although dormant for a few years, the intellectual agenda, open source code, and methodologies of *OpCit* remain relevant for researchers such as the Human Voices project, Gregory Crane at Tufts, CiteSeer, and the OAI initiative.

Quotation identification is a subset of citation identification that is particularly significant because quotations create a logic of semantic association between units. As Ernst and Crane (2008) point out, there is little existing work that automates matching and identifying quotation and allusion between primary texts of a historical nature in cases where authors don't already structurally identify such affiliations. Ernst and Crane propose an algorithm for identifying patterns in texts as likely candidates for hidden instances of quotation and reference; we would welcome the chance to adapt their algorithms to our project to work with the historical archives of the early modern period. Due to the vagaries of our datasets, though, this algorithmic approach would largely function as a quality control assurance, since the historical archives we use (EBBA and Spenser) consist of highly structured texts where citation, paraphrase, allusion, and quotation are well understood and already tagged by experts in the discipline. Nevertheless, Ernst and Crane demonstrate the feasibility of an automated approach to textual citation beyond mere parenthetical reference and metadata extraction.

Identifying and extracting citations is one thing; using the presence of citations to generate concept maps for domains of knowledge is a much more difficult affair. One difficulty is the non-standardization of references between scholarly papers and venues. Andrew McCallum has worked extensively in this area, conducting feasibility tests that identified the difficulties of canonicalization, abbreviation, and other ambiguities in text citations (2007). Goldstone (2004), Bradshaw (2002, 2001), and Hammond (2002, 2001) have developed approaches to citation that link citations to indexing based on keywords. None of these approaches, however, uses citation relationships to create detailed models of domain knowledge. Instead, existing approaches seem to suggest that citations themselves tell a user surprisingly little about the propositional content of an article, since there are disciplinary biases towards citing specific authors and recent texts.

More recently, Fuzzy Association Concept Mapping (FACM) has been suggested as a promising model for developing concepts in texts processed by Natural Language Processing abstraction. Wang, Lee, Cheung and Kwok (2008) develop an information organization model that recursively asks the user to interact with automated concept maps, refining the map through human interaction. Similarly, Ritchie, Robertson, and Teufel (2008) suggest indexing articles using the sentences that surround citation points. These researchers indicate the potential of the surrounding text of what we term "citation units" as of particular semantic value. However, by focusing on keyword extraction, their methodology fails to overcome the limits of current approaches to keyword mapping and searching.

Our environmental scan reveals a sufficient number of research teams working in close parallel to our methods to suggest that our approach is worthy of pursuit. No published study aligns with our goals, though, in providing a reconfigurable network of affiliation relationships to visualize concept maps in the discipline of the humanities. We are cautiously ambitious that our approach will yield results of interest to these researchers and others pursuing similar lines of inquiry.

## **Project Members**

Growing out of several projects in the digital humanities—MaPs, EBBA, Spenser Archive—the Human Voices project insists on a productive simultaneity between humanistic inquiry and computational data mining. The broad range of team members represents a sustained focus on delivering useful tools and methodologies that can serve researchers from a variety of fields in the humanities, social sciences, and

sciences. Our project team has little duplication, instead bringing together stakeholders into a jointly beneficial mutual endeavor.

The team at the National Center for Supercomputing Applications and UIUC, led by Michael Welge, brings a demonstrated interest in algorithmic data mining to yield results that are useful to humans. Their SEASR platform, an open-source and community centered environment for scholarly research that uses a web services model as a browser plugin, is under widespread adoption in the humanities for its powerful flexibility in text and data analysis. The automated learning group at NCSA, led by Loretta Auvil, is insistently collaborative, routinely partnering with external researchers to incubate and sustain leading-edge techniques in data mining. NCSA has access to staff with diverse expertise in areas critical to the development and support of projects such as ours, from visualization experts to researchers in applied math. The strong collaboration of the NCSA team is a critical component of the success of Human Voices.

The team at the University of South Carolina, led by Randall Cream, consists of a diverse set of expertise and a significant private-public partnership that represents a substantial investment in the Human Voices project. The Center for Digital Humanities at South Carolina is committed to sustaining and developing useful tools in the humanities; Human Voices represents an exciting opportunity to collaboratively coordinate efforts that exceed the scope and scale of research currently possible at the CDH. Randall Cream and Jun Zhou have combined on a wide array of projects in their brief time together—from a Humanities Gaming Institute and software to support digital editions to building useful tools to enable textual scholarship. Their work on Human Voices reflects a continuation of their ongoing interest in data mining as a stage in modeling cognitive behavior. The CDH's work with the Spenser Archive, led by David Miller, provides a mutually beneficial relationship for both the archive and the data mining initiative. Duncan Buell and Manton Matthews lead the Computer Science Engineering program at the U of SC, with current research interests in data mining, pattern recognition, and information retrieval. These CS researchers also bring an external partnership into the team with the software firm Collexis. Collexis builds software that allows predictive and anticipatory results building in the biomedical sciences, generating recursion effortlessly as a component linking inquiry, writing, and researching. We're proud of a strong and beneficial partnership that offers us a proven platform to model our results through, and the chance to learn from one of the leading vendors in the marketplace.

The team at the University of California-Santa Barbara, led by Patricia Fumerton, brings both an amazingly diverse and valuable archive of broadside ballads, EBBA, and demonstrated experience in successfully data mining popular humanities texts that speak across disciplines (as art, literature, history, and music). Eric Nebeker and Kris McAbee offer project management experience and the ability to develop interfaces for digital work in the humanities that humanities scholars actually use. Their EBBA project, multiply funded through the NEH, routinely wins awards and generates scholarship interrogating the space for digital inquiry in traditional humanities disciplines. Another key participant in the UCSB team is Carl Stahmer, a researcher in data mining and visualization interfaces in the humanities for over twenty years whose work includes the very successful NORA and Monk project. The UCSB EBBA team has considerable experience successfully navigating the difficulties of large-scale distributed projects, and brings that awareness to Human Voices.

The team at McGill University, led by Paul Yachnin, brings a diverse set of experiences in yielding revolutionary results using digital methods in the humanities. The McGill researchers belong to the SSHRC MCRI 'Making Publics' (MaPs), an international research project examining the formation of informal groups of association and collaboration in Early Modern Europe. As part of its mandate MaPs has undertaken exploration of large-scale international collaboration within the humanities around a given research thematic. Its partners bring considerable experience to bear in regards not only to conducting humanities collaboration, but also the study of ideas and the formation of groups around works of art,

literature, science, and academia. Rather than bringing an archive of scholarship to the project, our partners offer their experience in building and designing interactive recursive environments for scholarship. Matthew Milner has led the move as part of his interest in social networking models for humanities research and the federation of thematic-focused archival resources. Renee Sieber, working in Geography, the School of the Environment, and Computer Science, has led several projects that demarcate the line between the social sciences and the humanities, applying computational methods to humanities interpretations to answer urgent questions about affiliation and space. Jean Guy Meunier, from Université du Québec à Montréal, has engaged in a variety of semantic approaches to text mining, using artificial intelligence to build fuzzy recursive systems that self-refine. These core researchers, deeply experienced on multi-site collaboration with the MaPs project, also bring a noteworthy external collaboration into the project. Vlad Petric, a software engineer at Google, collaborates with the McGill team on search design, information extraction, and using metadata to build interactive user applications.

These four teams are separately skilled with each having a unique ability to combine computational methods in the humanities with original humanistic inquiry. Together, the four teams represent a concentration of excellence that far exceeds the scope of the Digging into Data project. By building this impressive team, we hope to demonstrate our commitment to delivering results within the timeframe of the grant (January 2010 to March 2011) and beginning a project that grows beyond the contours of the initial investment.

### **Project Scope & Duration**

The large scale of the Human Voices project necessitates a distributed project that coordinates multiple simultaneous tasks in order to allow efficiency and collaboration to quickly develop amongst the partners. The aims of the Human Voices project, to deliver a flexible, modular data mining application that scholars working in the humanities will actually want to use, require coordination on several simultaneous goals. We envision our project spanning the entire grant period, January 2010 to March 2011.

First of all, the project must write the code that performs the analytic routines in Phase I (citation identification), Phase II (citation unit segmentation by associating contextual sentences with the citational relationships), and Phase III (aggregation and association of citational units to generate concept maps and facilitate semantic analysis of disciplinary concepts). This sort of process is inherently recursive—it uses the output of one stage as an input for another stage of analysis that passes over the same data, creating a workflow that is largely sequential. The UCSB team will take the lead in adapting and porting the open-source code generated for the NORA and Monk projects to our more specific purposes. With Carl Stahmer's firsthand experience on the NORA project, we feel confident that the UCSB team can adapt the material to new purposes in a timely fashion. The SEASR team at the NCSA/UIUC will work to ensure that the SEASR platform can reliably sustain the computational load of the data mining proposed by Human Voices, supporting the growth and development of the project in the collaborative environment of SEASR. The team at SC will write new modules for SEASR, using a variety of languages including python, Java, UIMA, and other frameworks, refining and developing computational methods for citation identification, extraction, and disambiguation. The team at McGill will develop an interactive and visualization layer for each phase of the work, allowing scholars to interact with the processes and visualize the data to build textual maps. Loretta Auvil at NCSA, Randall Cream at SC, and Kris McAbee at UCSB will work to coordinate visualization and interaction layers with the underlying data mining.

Given the dependencies involved, we recognize that the first phase of our project (automated citation identification and extraction) constitutes a bottleneck of sorts on the project. We anticipate this crucial stage of code development to last about ten months for each team, with the application that provides the visualization and interaction layers trailing by approximately three months. To create efficiencies, we'll work through the well-structured early modern archives that already contain citation tags, essentially

seeding our environment with a few dense relationships. This dual-stage work over the first phase should allow project staff the time to recursively refine the algorithms to address the peculiarities of the JSTOR data model efficiently. On the other hand, the second phase, segmenting the texts using this extracted citation data, is far more transformative in terms of human-computer interaction, but much easier computationally. During the remaining five months of the grant period, we anticipate completing the second stage or making enormous headway to completion. To make sure the collaboration stays current over the course of the grant period, Human Voices project will meet twice during the fifteen months, facilitating a means of interaction that can help constitute an effective international partnership that can survive the vagaries of a 15 month project.

From our timelines, the ambition of our third phase—semantic analysis and concept mapping between texts—becomes apparent. We anticipate the work of Human Voices project to continue well beyond the fifteen month frame. Specifically, the semantic analysis and concept mapping modules are computationally much more intensive than analysis of citations and segmentation along instances of quotation. Our project teams will likely require the entire grant period just to develop the first two phases of the project, with the associated visualization and interaction layers that make such tasks worthwhile in the humanities. Rather than concentrating on performing as much data mining as possible within the grant window, we'll work to ensure that mature tools are disseminated, usability is built into these tools, and the project works from the outset with an eye toward structuring and maintaining a collaborative environment for research.

Our teams bring disparate goals into collaboration in order to tie us together in an effective partnership. Some teams are interested in maturing a software model that has already proven itself competent and successful; others, such as JSTOR, are interested in delivering innovative new tools that have persuasive utility to analyze culturally significant material. In essence, this DiD challenge will serve as a startup incentive to form a stable coalition, balancing a host of concerns against an array of possibilities. We anticipate a long-term collaboration, and have built in two public-private relationships (Vlad Petric, software engineer at Google, and Stephen Leicht, COO at Collexis) that hold potential for our project after the grant period is complete. Human Voices will seek a variety of funding models to sustain its effort after the grant period, including internal funds, entrepreneurial partnerships, and government sponsored funding opportunities.

### **Anticipated Outcomes**

The Human Voices project undertakes three primary outcomes from its fifteen month grant period. First of all, the Human Voices team seeks to develop and sustain a collaborative environment for data mining as a human-centered act, an act which develops out of the participatory and interconnected nature of human culture. Our project invests heavily in the collaborative environment of SEASR, developing interactive components to ensure it remains relevant for humanities researchers. Secondly, we want to return significant value to the archives which contribute their data to the challenge, by developing secure, mature, and distributable tools that maximize the utility and significance of the humanities data within their archives. And thirdly, our project conducts innovative and worthwhile research into data mining in the humanities, creating a model for automated citation identification and extraction that should assist researchers in a variety of fields and disciplines.

Publicity is an important element of attending to the responsibilities of the public trust reflected in competitions such as Digging into Data, and Human Voices will actively work to ensure that there is a large impact of the substantial governmental investment in our project. With such a large and distributed project, word of mouth and social networking at conferences and discipline specific events is an effective strategy for disseminating project results. Of course, our project will invest in building a white paper for

the results of our data mining research, and heartily participate in the Digging into Data conference to discuss approaches to large scale data in the humanities.

The real impact, though, of Human Voices will be a sustained collaborative partnership that crosses national borders and disciplinary barriers to construct a focused and sustained team for data mining in the humanities. We've worked diligently to recruit private partners to share in this public investment, suggesting that our project enjoys a relevance not only to university researchers but also to archivists, information workers, and anyone interested in mechanisms which bring the massive data from the past into useful focus for individuals far removed in space and time from the creation of that data.

### **Budget Overview**

Human Voices consists of four teams of researchers across North America (McGill, UCSB, NCSA, U of SC) and two significant private partnerships with leading-edge companies in information science (Google and Collexis). Our fifteen month project invests almost the entirety of its funds— 100,000 dollars from each of SSHRC, NEH, and NSF—in labor, the limiting element in most projects in digital humanities. Our approach to the Digging into Data challenge is to treat this competition as a lengthy start-up project, where public monies entice an investment by private corporations and experienced researchers. Given the scales of the problems we tackle, and the dimensions of the data we have chosen to work with, we do not enter into such a partnership lightly. Instead, our commitment to this endeavour is reflected in the fact that each of the four teams commits its leaders almost entirely pro bono, choosing instead to fund graduate students, young researchers, and junior career personnel. We have joined together to form a working group with an understanding that the best results may resist us for much of the grant period. An investment in Human Voices is an investment in the persistent forms of collaboration that facilitate research programs such as our own.



**Stibos Project**  
**Modeling Information Research Behavior as a Component of Scientific Inquiry in**  
**Undergraduate Education**

**Competition: NSF Cyberenabled Discovery and Innovation, From Data to Knowledge**

**Submitted: December 18, 2008**

**Proposed Award: \$1,400,000**  
**Unfunded**

**stibos.org**

Randall Cream, PI  
Christy Friend, Duncan Buell, Caroline Eastman, Co-PIs.

Stibos project is a large-scale, longitudinal study of the relationship between research, citation, and inquiry in undergraduate education. Envisioned as an element of information literacy curriculum revision with the Director of First-Year English as a Co-PI, I designed a plug-in to Zotero to readily sync Zotero data stores with a central server. Guarding anonymity and personal data, we ran a pilot of the program in the Spring of 2009. Using the results of that pilot group, we refined the software and are launching a larger scale in the Spring of 2010. Eventually, we'll compile the largest archive of student research—not reported behavior about research, but the actual micro-decisions that shape and guide inquiry as it occurs in undergraduate research.

Stibos was not funded in its initial phase, but we've gathered enough data to prove our prototype works. Our reviewers suggested we separate the different aspects of the project and solicit funding separately for the various dimensions: data collection, curriculum revision, and data mining. We're currently pursuing the first phase of funding for data collection.

**CDI II: Modeling Information Research Behavior as a Component of Scientific Inquiry in Undergraduate Education**  
**PI: Randall Cream, University of South Carolina**  
**From Data to Knowledge**

Our project offers the first large-scale longitudinal study of undergraduate student information-gathering as a predictor of sustainable research behavior. Our project contains three separate and connected processes: First, using a widely-available open source software plug-in for the Firefox browser, our team of researchers will deploy a campus-wide information literacy initiative to train students in information gathering as a core skill for inquiry, beginning in the required first-year composition and research courses and sustained as those students move throughout the colleges and departments of the university. Second, our research team has developed software for this browser plug-in, Zotero, which allows accurate, complete snapshots of ongoing student information research behavior to be gathered using a central server. Third, our team will aggregate and mine this data store, indexing the heterogeneous results using Latent Semantic Analysis through Singular Value Decomposition to reduce the dimensionality of the data and create useful models of student information research behavior as a correlative of educational performance, within the sciences and across the university.

**Intellectual Merit:** Despite a sustained interest in student information research behavior, there is a startling lack of large-scale longitudinal studies of this behavior (Williams and Rowlands, 2007). Even relatively well-gathered studies of student information research have become hopelessly outdated by the rapid transformation of the internet into a credible and sustainable resource for students to interact with journal articles, scientific papers and studies (D’Esposito and Gardner, 1999). As information research gathering has quickly shifted from using the computer to locate offline sources (databases) to using the browser to interact with those sources directly (digitized information resources), our ability to meaningfully analyze and accurately model student research behavior has fallen further and further behind. Even while our colleagues have worked to explore the utility of internet-based peer review as a component of scientific inquiry (Timmerman et al, 2003), and recent studies have suggested that upwards of “forty million Americans rely on the internet as their primary source for news and information about science” (Horrigan, 2006), these studies nevertheless remain limited to *reported* behavior, rather than verifiable representational data about information research gathering practices. Our project will produce the first-ever independent assessment of student information research collection, longitudinally tracking a student base of approximately 3500 students a year (in the first year) throughout their undergraduate career, compounded with each incoming class to total approximately 14,000 students in year four.

**Broader Impacts:** More broadly, the impact of our project lies in its ability to facilitate questions interrogating the relationship between student research behavior and patterns in information sources. What relationships can we identify between student performance and a semantic analysis of the research conducted by those students? What features in the data conform to cognitive choices students make in selecting sources for their inquiries? By relying on snapshots of *ongoing* research behavior, rather than self-reporting or analyses of *completed* research, we can gain an understanding of student inquiry as it takes shape, and not just in its finished form. Such an understanding is indispensable to any systematic attempt to influence undergraduate education in the sciences and beyond.

## **List of Participants**

### **CDI II: Modeling Information Research Behavior as an Indicator of Scientific Inquiry in Undergraduate Education**

#### **Principal Investigator:**

Cream, Randall. Center for Digital Humanities/English Department, University of South Carolina.

#### **Co-Principal Investigators:**

Buell, Duncan. Department of Computer Science and Engineering, University of South Carolina.

Eastman, Caroline. Department of Computer Science and Engineering, University of South Carolina.

Friend, Christina. First-Year English/Department of English, University of South Carolina.

#### **Key Collaborators:**

Zhou, Jun. Center for Digital Humanities, College of Arts and Sciences, University of South Carolina.

## **CDI II: Modeling Information Research Behavior as an Indicator of Scientific Inquiry in Undergraduate Education**

**PI: Randall Cream, University of South Carolina**

**Co-PIs: Duncan Buell, Caroline Eastman, Christy Friend, University of SC  
From Data to Knowledge**

We know that transformative knowledge construction is strongly associated with students learning to structure problems efficiently and productively (Shin et al, 2003). This knowledge is not acquired in isolation, though; just as we do, students build scientific inquiry by developing an ability to “sort out, evaluate, and organize their diverse views” about conflicting evidence (Linn, 2003). Researchers have concluded that students are often unable to pose meaningful problems without first attempting to construct what Linn terms a “knowledge integration perspective” facilitated by recursively considering a concept from a variety of perspectives (2003; also, Novak, 2002; Pelaez, 2002; Johnson & Johnson, 1999). While information research—drawn from original sources; published, expert-led studies; and white papers, case studies, and other documents—offers the raw materials for students to assemble this self-multiplying perspective, past studies have found that students are often unable to build the “compelling comparisons” (Davis, Bell & Linn, 2003) necessary to make sense of information research integration. These researchers have suggested the utility of an investment in “language literacy” embedded within “science literacy” (Linn, 2003), where scientific inquiry is predicated on an ability to combine, compare, and disentangle competing expert opinions (Rivard & Straw, 2000). Other studies have suggested that this inability to construct a multifaceted approach to problem building leads students to dismiss science education as irrelevant, suggesting that the impediment is not one of difficulty and abstraction but rather one of situatedness and resource/research literacy (Stocklmayer & Gilbert, 2002; Linn & Hsi, 2000; Kintsch, 1998; Songer & Linn, 1991). For science education to be successful, it must evoke students’ ability to not only answer complex questions, but also the ability to ask those questions in ways that draw upon, adapt, and negotiate with the work of others (Linn, 2003; Bell & Linn, 2002; Pelaez, 2002; Zohar & Nemet, 2002). While there is extensive research that suggests the critical importance of the skill of evaluating, interpreting, and integrating the knowledge that is produced from the data of information research in scientific inquiry, there is to date no study that examines these skills *in vivo*, as an active cognitive process instead of a product of consciousness.

Our research group aims to produce the first-ever study of student information research behavior as a process of inquiry, drawn not from an analysis of surveys or the finished results of problem-solving, but rather from an investigation into the process of information searching as it occurs. We are interested in studying not just the *results* of information research but also the deliberative process of that research in order to understand the micro-processes of information juxtaposition, opinion weighting, and hypothesis generation as components of inquiry. In short, we are just as interested in the perspectives that a researcher considers useful but ultimately rejects as we are in the actual studies and reports incorporated into the finished papers and projects. Rather than mine the results of students’ papers and reports, we propose to use software to maintain an accurate snapshot of student research as it is conducted, enabling us to reconstruct the cognitive dimension of research and unfold the richly complex and subtle steps of knowledge building. Our team of researchers—drawn from engineering, the sciences, the arts, and computer science and information technology—has constructed a project that is flexible enough to sustain a four-

year integration into the general education curriculum and sophisticated enough to model individual and sub-group behavior within that curriculum. In the course of implementing a longitudinal study of four-years of student information research activity, we'll generate a data set that is immensely valuable to disciplines as diverse as information science, psychology, sociology, and computer science, not to mention its utility for constructing accurate models of student research behavior in science and undergraduate education. If the goal of science education, if not education itself, is to enable its practitioners to ask meaningful questions about relevant problems, and we accept that the very ability to ask those questions is constrained by our fluid navigation of information resources and data collection, then it is increasingly imperative to be able to accurately model the ways that students navigate this difficult terrain.

The innovation of our project, therefore, lies not in its desire to better understand student information research activities as a component of scientific inquiry, but rather in its ability to extend recent trends in web technology to provide the capability for accurately capturing the sub-routines of information retrieval as they occur for each student. The open-source software tool, Zotero, provides our project with a stable, widely-recognized tool for the basic function of information organization. As a plug-in to the Firefox web browser, it extends a familiar interface for information research and retrieval. Zotero works by cataloguing research brought up on web pages using the embedded metadata in page headers, but it also automatically downloads and stores linked full-text pages and attachments. The software team at the Center for Digital Humanities at the University of South Carolina, led by Jun Zhou and Randall Cream, has developed an extension to the Zotero plug-in that allows a remote sync of the data store to a unique directory on a secure server. Operating independent of user intervention, as students conduct their research and decide to add sources to their Zotero bibliographies, our extension syncs the activity automatically and securely. Over the course of semesters and years, each student's data store grows more and more complex, coming to represent a multi-faceted picture of their intellectual growth and development. The data store, anonymous but uniquely tracked, at the end of four years contains information that is almost priceless in its ability to represent knowledge-creation at the microscopic level. As such, our project deploys computational thinking to leverage bibliography-construction into information research literacy.

More pressing, however, than its ability to analyze information organization to construct a model of scientific inquiry, is our project's plan to deploy that computational thinking as a direct pedagogical device. On a curricular level, our team (led by Christy Friend and Randall Cream) will institute an implementation of Zotero that begins in first-year research and composition courses (offered within the English department) as a component of the University's information literacy requirements. In those courses, directly supervised by Friend, students will receive a flash-drive imaged to contain a portable installation of Firefox with Zotero that will launch in Windows and Macintosh environments. With approximately 90% of the first-year student population as clients, these gen-ed required courses are appropriate units to deploy our information literacy campaign. With an average population count of 22 students per class, the first-year sequence is also uniquely able to offer students a level of personal assistance uncommon in large-scale projects such as this. Through our research project, students will learn not only to use the Zotero tool to organize and retain the results of their research, but will also learn to think of the behavior of information research computationally. In learning to conduct information research as a component of problem solving, students will be informed of our

research project and have the opportunity to give consent to the use of their research data. Reinforced by teaching assistants working under the training of Friend and Cream, first-year students will receive regular reminders that their data is being captured, analyzed, and modeled. In this way students will come to think of their own behavior as modelable and analyzable. As a result of this research project, students will learn to apply computational thinking to their own activities as a condition of conducting information research. By studying information research behavior, we are able not only to turn our students' attention to the important skills of information literacy as a component of problem solving, but we also are able to induce and encourage the development of computational thinking as a human behavior.

While one significant component of the innovation proposed by our research project institutes a model of information research as both a subject and an avenue for computational thinking, we are nonetheless equally excited at the consequential opportunities for creating analyses and models from our data set. Our research project will yield unprecedented access into the deliberative process of information gathering as students negotiate between conflicting opinions and perspectives to build their own research questions. With Zotero's inherent ability to effortlessly capture data from web-based research, we predict that students' research data store will greatly exceed the number of items actually used in the production of finished papers, projects and reports. With the help of Friend and Cream, we intend to train students to add much or all of the data from their deliberations to their data stores, encouraging them to use information research as a tool in inquiry, not just as a support for already-conceived opinions and evaluations. As Zotero catalogues the metadata, takes snapshots of the pages, and archives the linked attachments, each student's data store takes shape as a reflection of that student's process of inquiry. As a final item in the process, our graduate teaching assistants will encourage students to add the finished products of their research into the Zotero store, allowing us to archive an idea throughout its many iterations and variations, including its final form. The data set for our research project, then, will begin with approximately 10,000 separate but complete deliberative processes a semester (3 research assignments x 3500 students), representing each student's research work through the first-year sequence of required research and composition courses. Each process of inquiry yields important data about student decision making, information organization and retrieval, and assessment. Together, the 21,000 separate information research projects each year will serve to transform our understanding of how students think through the research problems they encounter at our university.

In addition to collecting data on each year's first-year students, we are also interested in investigating the relationship between research and inquiry longitudinally for students over the course of their undergraduate career. While we anticipate some drop-off in student involvement from semester to semester, we'll undoubtedly retain enough involvement to gather meaningful data about a time-value decomposition of research behavior across student populations. Our research team has forged links to most of the majors in arts, sciences, engineering, and education, ensuring the uniformity of information research and retrieval methodologies across campus. Assuming an attrition rate of 30% per year, our project would nonetheless produce meaningful, four-year data for approximately 1200 students a year. It would be difficult to over-emphasize the transformative possibility of this data in an area where almost all the studies are conducted by survey, and the largest observational study with longitudinal results involved 188 subjects (Williams and Rowlands, 2007). Using algorithmic analysis, including LSI through

SVD and the more demanding Genetic Analysis routine suggested by Song and Park (2008), our research project will be able to link student decisions made during research gathering to individual and group performance outcomes on a granular level. We remain especially interested in determining links between types of research behavior and demographic subgroups, including ethnicity, college major, gender, and GPA.

Even as Cream, Friend, and Zhou work to supervise and train the Zotero implementation, data collection, and data archiving aspects of the project, Buell and Eastman will work with graduate research assistants in Computer Science and Engineering to develop efficient methods for structural and semantic analyses of the Zotero data. While we propose several logarithmic methods for decompositional analyses of information research data, we anticipate that the richness and diversity of the data will suggest adaptations to existing algorithmic models, and may even dictate the creation of a new algorithmic model for analysis. Eastman and Buell will work with the graduate student as the data is collected, in real time, supervising the graduate students as they work through the large data set that accumulates.

The framework of our team—with Friend’s expertise in composition, assessment and information literacy; Cream’s expertise in applying computational methods to questions in the humanities; Eastman’s expertise in information organization and retrieval in natural language systems; Zhou’s expertise in building software tools across multiple computing platforms; and Buell’s expertise in number theory and algorithmic computation—constitutes an intellectual partnership that values diversity more than disciplinary. Uniting our disparate concerns and fields of inquiry, however, is an unyielding interest in effective undergraduate education, determined by students’ problem-solving and problem-posing abilities. Together, our team understands that our true education agenda is overcoming the challenge of viewing education as an object-oriented existence, where students come to see the world as composed of discrete fields and disconnected objects and problems. Each educational activity, in this view, is a separate and self-contained struggle. Our research goals, however, are to develop cross-disciplinary tools that affect the ways students understand themselves in relation to the problems posed in education. Thinking about themselves computationally—understanding their actions as recursive process that iterate across common objects—is a significant step towards that end.

**Handheld Art: Mobile Devices and Participatory Learning Technologies  
in the Secondary Art Classroom**

**Competition: NEH Summer Seminars and Institutes for High-School Educators**

**Submitted: March 3, 2009**

**Proposed Award: \$140,000**

**Unfunded**

**[cdh.sc.edu/handheldart](http://cdh.sc.edu/handheldart)**

Karen Heid, PI

Randall Cream, Co-PI

Karen, a faculty member in Art Education, began discussions with me in Spring 2008 regarding the NEH summer seminars. With the March deadline, we began working on a March 2009 deadline and a 2010 seminar. Initially interested in seminars that updated technology in art education, we quickly settled on the mobile phone platform as an innovative approach to content delivery. For this application, Karen provided the educational framework and the expertise in secondary education, while I devised a seminar that would introduce educators to the aesthetics and interface constraints of the mobile platform. I wrote this application, while Karen recruited the experts in art education to anchor the project. I promised the resources of the CDH to assist local educators in server-based content for the handheld platform.

This project was not funded. While four of the five reviewers were VG or E on the reviews, one astute observer noticed that there was a tension between studying art and making art, which is unfundable under the NEH guidelines. This position introduced enough questions to lead the other reviewers to lower their initial reviews substantially in the final round, resulting in an unfavorable decision.

We will resubmit with a tighter focus in the next competition.



**Handheld Art: Mobile Devices and Participatory Learning Technologies  
in the Secondary Art Classroom**

Summer Seminar for High School Teachers July 12-30, 2010

Karen Heid, Project Director,

Co-Directors Heidi Rae Cooley, Randall Cream, and Simon Tarr

University of South Carolina

The proliferation of mobile technologies is reshaping our world in ways that are only partly visible, as we rapidly move beyond industrial models of economy to a society built on networks of information exchange. As early as 1988, Mark Weiser argued that the network was producing a massive shift in human culture even as computing itself morphed from the framework of clients and servers to metaphors of desktops and laptops. Weiser suggested that in the near future, computational capabilities would be so pervasive that human culture would participate in what he termed “ubiquitous computing.” Thoroughly integrated within the infinitely varying manifestations of human activity, the assimilation of human and machine into an indistinguishable and interwoven fabric of existence has profound implications for those interested in the production, manipulation, and interpretation of human culture. Handheld mobile devices, easily the most ubiquitous computing platform on the planet, present a unique opportunity and a persistent challenge to existing models of cultural transmission; nowhere are these challenges more visible than in the classrooms that shape and channel the efforts of teens and youths to move from receptors of cultural information to producers of culture.

Mobile phones, as one of the more noticeable conduits for the networked flows of information that comprise social and political participation, have a somewhat tenuous existence

in secondary classrooms. Most districts continue to discourage mobile devices on school campuses, even as researchers have begun to underscore the educational possibilities contained within the mobile phone platform as a pervasive computing device. In light of recent initiatives in the UK ([mLearning Project](#), [Mobile Learning Network](#)) and in the US ([Mobile Learning 2009](#), [Educause Horizon 2009](#)), we believe the time is ripe to support sustainable projects that develop ubiquitous computing technologies in secondary education. We propose a three-week seminar for secondary teachers of art, inviting them to discover, adapt, and invent new and existing methods of teaching in response to evolving media technology and participatory culture.

Our team of researchers, with expertise in art education, new media arts, teacher preparation, mobile phone culture, and digital humanities, proposes to lead a three-week seminar to help secondary art teachers investigate a pedagogy of the mobile phone platform and develop a participatory learning environment through the use of handheld devices in secondary art education. Led by Dr. Karen Heid, an experienced scholar of multiage art education (see vita, page 20), our project will engage participants in thinking through practical applications for a pedagogy of mobile phones in the art classroom. With Dr. Heidi Rae Cooley, participants will investigate the aesthetics of the mobile platform—examining the productive limits of the platform’s interface, tracing the complex ways that handheld devices mediate social interaction, and exploring the processes by which mobile devices juxtapose computational thinking and the poetics [*poiesis*] of the everyday (see vita, page 25). Working with Dr. Randall Cream of the Center for Digital Humanities, our participants will develop practices for implementing the mobile phone as teaching tool in a sustainable way, utilizing the university’s investment in cyberinfrastructure to decrease the initial barriers to adoption and provide a model for sustainability that does not burden the infrastructure of the participants’ home schools (see vita,

page 30). In sessions led by Professor Simon Tarr, the seminar will work to consider mobile culture as a platform for video, moving images, and still photography. With consider expertise in the aesthetics of the screen and the multiple affordances of the mobile platform, incorporating recent innovations such as game-based learning environments, Tarr will provide seminar participants the chance to interrogate recent work at the forefront of new media. With guest lectures from leading experts in new media technologies—Stephen Carpenter, Robert Quinn, and Pamela Taylor are emerging as some of the nation’s leading young voices in adapting art education to the unfolding technologies of visual culture—and a core of project faculty with complementary expertise in art education, teacher training, new media art, and humanities cyberinfrastructure, our seminar is well-positioned to support a diverse group of participants working quickly to develop familiarity and proficiency in this transformative new platform for participatory learning (see Project Faculty and Staff, page 10).

### **Intellectual Rationale**

As computational technology matures, we are entering a phase where innovation emerges not just in the invention of never-before-seen devices, but also in our imaginative restructuring of familiar devices to produce suddenly-new reorganizations of daily routine. We are in the midst of just such a moment, as the ubiquitous cell phone is metamorphosing into a compact, mobile, networked computing device. As early as 2003, researchers noticed a rapid decline in voice-based communications as the predominate usage of cellular network bandwidth, signaling the shift of the device from a phone that connects via voice to a computing platform where users interact not with others but instead with an extensible interface. Three key features of the mobile phone capture its unique capability as an emerging platform for educational computing. In the first place, the mobile phone is ubiquitous; Forrester Research asserts “market saturation” for

mobile phones as cell phones ownership rates amongst late teens approach nearly 100% of the high school population, representing a greater penetration than either the desktop computer or cable television. For teachers and administrators who have spent years addressing the “digital divide,” the mobile platform achieves what years and billions of dollars of infrastructure investment could not: a reliable last-mile strategy. Secondly, the mobile phone is extremely portable, as part of a design carryover from its earlier manifestation as a voice platform; this portability, both in terms of battery life and physical size/weight, means that students routinely carry these devices with them and use them almost constantly. For teachers and administrators interested in uniting students’ learning across multiple social spaces, both within and beyond the classroom, mobile phones represent the best chance to facilitate asynchronous participation for students who have learned to adapt to an ever-growing array of responsibilities—from work and school to extra curricular activities and sibling proto-parenting. And thirdly, the mobile phone is a curiously *personal* device; for reasons and in ways that researchers have only just begun to study, individuals experience an intimate relationship with the mobile phone. As project co-director Cooley argues in “The Autobiographical Impulse and the Aesthetics of the Mobile Phone,” the mobile phone suggests to its users that they take the device as an extension of themselves, in their hands, internalizing and sublimating its interface as a natural extension of self-expression. Even as teachers and administrators bemoan the insertion of the syntax and vocabulary of texting into academic spaces, we miss a far greater possibility signified by the presence of this intimate computational device: a powerful assimilation of learning and self that has the capacity to meet traditional educational models using new media methods. Responsibly capitalizing on these features of the mobile device platform presents a significant challenge to educators in the 21<sup>st</sup> century.

The time is ripe for large-scale adoption of mobile computing as pedagogical activity. The 2009 Horizon Project, whose annual forecast of computing in education has become a touchstone for educators and technologists alike, compellingly urges that the “capabilities of phones have continued to develop rapidly” enough to place “mobiles firmly in the near-term horizon,” a category referring to “emerging technologies likely to have a large impact on teaching, learning, research, or creative expression within learning-focused organizations.” Sidestepping issues of classroom management, the report points out that “it is apparent that the devices and their new applications have been accepted in the mainstream” of society (Horizon 2009). Part of the resistance to mobile platforms in the classroom stems from their inherently personal nature; with each student able to work independently on a computing device that is both connected and private, networked with others and also viewable only by its user, the underlying model of education as a disciplinary means of control is thereby exacerbated. However, if more than half the world’s population connects to the internet from a mobile device by 2020, as predicted by the most recent Pew Internet & American Life report, then our resistance to the handheld device as disruptive to teaching and learning seems even more fruitless. In fact, such a degree of saturation would make any pedagogy that fails to attend to and incorporate these devices itself disruptive. As a “universal tool for communication of all kinds” (Horizon 2009), we should recognize and be responsible to the indisputable fact that we are witnessing the emergence of a dominant medium, one which can facilitate the production and study of culture in a dizzying array of forms.

Proposing that mobile phones serve a viable ubiquitous computing application in classroom instruction addresses a growing discrepancy between the skills and abilities we value—critical reading and thinking, literacy, creative interpretation, and knowledge

integration—and our inability to translate these skills into the media that represent most students’ engagement with information. Information literacy, visual literacy, and technological literacy remain poorly integrated in the traditional subject areas, while auxiliary areas such as art and music are so challenged by budget cuts and underfunding that they are unable to respond to the shifting terrain of their fields. Our seminar aims to respond to that disparity by preparing its participants to work as catalysts of change within their departments, schools, and districts. Art is especially well-suited to this catalytic role, since it has grown accustomed to the experimentation and stochastic meddling of politicians and parents for several decades. Informed by the recent Visual Culture arts education movement, our seminar proposes a necessary corrective to the entrenched methods of teaching art as a discipline-based curriculum where students engage in art production, art history, aesthetics, and art criticism as more-or-less discrete components (Anderson and Milbrandt 2005). Just as we’ve learned to understand the ability to read books as a determinative skill for success in areas other than the field of English, we argue that visual literacy—the skill of working with information embedded in and through visual media—is constitutive of other disciplines and is equally determinative of success in education. Our seminar encourages participants to build on the strategic advantages inherent in the mobile device—its ubiquity, its portability, and its intimate connection with the individual—to develop a pedagogy for students to learn to think in the language of images as a critical engagement with their environment.

### **Seminar Content**

Our three week seminar is designed to prepare its participants to work with emerging media and handheld content in the secondary art classroom. The seminar’s content works recursively between theory and practice, simultaneously building both an awareness of current

debates in the field and proficiency in pedagogical uses of handheld devices. Along the way, we'll read articles and excerpts from books that highlight the rich possibilities inherent in participatory learning in and beyond the classroom. Project faculty from art education, new media technologies, digital humanities, and media arts will focus the discussions in specific ways on areas of local concern. Our guest faculty, drawn from leading voices in the field, will inspire and connect participants with pedagogical practices that are underway across the nation. Our detailed schedule (Appendix I) provides a full list of the work of the seminar, detailing individual responsibilities for our guest lecturers.

In the first week, participants will engage in a sustained discussion that considers the pedagogy and culture of mobile devices. Led by Karen Heid, art education, and Heidi Rae Cooley, mobile technology specialist, participants will simultaneously undertake a theoretical and practical examination of the mobile platform. Dr. Heid will lead discussions focused on the practical implementation of a curriculum that embraces the mobile screen. With readings from Elizabeth Hartnell-Young, Howard Rheingold, and Mitzuki Ito, participants will evaluate recent analogous developments in visual arts education and synthesize an understanding of the practical limitations for curricular change. Drs. Cooley and Cream will lead discussions focused on the culture and theory of handheld devices. With readings from Anne Friedberg, Vilem Flusser, and Manuel Castells, participants will consider the intellectual terrain of handheld culture, with discussions focused on synthesizing a coherent perspective from the multiple voices engaged in mobile culture and ubiquitous computing. Prof. Simon Tarr will lead discussions of the aesthetics of the mobile screen, investigating the practical implications of working with handheld devices and networked content. In the evenings and early mornings, the seminar will provide ample time for hands on work in the media labs and networked classrooms devoted to the

seminar. Our graduate assistant and undergraduate aid will engage with the participants in informal one-on-one meetings to help build proficiency and competency in the platform. On Tuesday, Wednesday, and Thursday, participants will engage in lectures and discussions led by Dr. Robert Quinn, a specialist in art education and distance education. Dr. Quinn's discussions will encourage participants to extend the classroom into other spaces, facilitating and encouraging synchronous and asynchronous activities outside the classroom. Throughout the week, Dr. Cream of the Center for Digital Humanities at U of SC will help work through the practical challenges posed as our participants work toward building a curriculum of the handheld device.

In the second week, our seminar moves from theory to practice, centering on questions of participatory learning through and with the handheld. Our faculty, led by Dr. Cooley this week, will suggest ways to undertake art education as citizen building. With readings from Walter Benjamin, Pamela Taylor, and Fujimoto Kato, participants will discuss the use of visual images to engage students in processes of knowing and discovering. Prof. Tarr will engage the participants in discussions and practice in teaching methods of composing for the handheld device. Drs. Cream and Heid will lead the participants in investigating podcasts as mobile phenomena, investigating mixed-media podcasts as narratives of environmental engagement. On Tuesday, Wednesday, and Thursday, Dr. Pamela Taylor, Chair of the School of the Arts at VCU, will lecture and lead discussions on the visual aesthetics of projection using mobile devices. Dr. Taylor will also lead our seminar in considering the ways in which the aesthetics of the medium mediate and alter content designed for the desktop. Continuing our commitment to mixing theory and practice, our graduate assistant and undergraduate students will help facilitate a hands-on approach to the technologies discussed in afternoon, evening, and morning sessions in



the media labs and classrooms. On Saturday, Drs. Cream, Cooley, and Tarr will lead an optional mini-workshop on writing applications for the iphone and ipod touch.

The third week's schedule moves from application to synthesis, with its goal being to assist participants in building three discrete lesson plans for handheld devices. Dr. Heid will lead the group through discussions of several model lesson plans built by our project faculty. Dr. Stephen Carpenter, Professor of Art Education at Texas A & M, will lead the seminar's discussions of participatory education as a feature of ubiquitous computing on Tuesday, Wednesday, and Thursday. With readings from Dr. Carpenter's work on visual culture and the civics of art education, participants will work in small teams and individually to build three lesson plans each. We'll spend the afternoons on Wednesday, Thursday, and Friday in workshop mode, facilitating an increased level of contact between seminar participants and project faculty. With three co-directors, a graduate assistant, and four undergraduate students assigned to the seminar, we'll be able to offer participants a chance to build the plans in a one-on-one environment. On Thursday and Friday, we'll share the lesson plans generated from the previous day. At the end of the seminar, participants will leave with a usb thumbdrive containing all of the project work, including the lesson plans from each member, the diverse readings, and technical literature and whitepapers. We'll mount all of this information on a seminar website hosted by the Center for Digital Humanities at SC, forging a permanent site to encourage continued collaboration and participation.

With a mixture of formal lectures, active discussions, social encounters at lunches, and workshops in the evenings and weekends, the seminar offers space for a variety of learning approaches. Our goal is to constitute, at the end of the three weeks, a meaningful core of educator-leaders who can lead change in their home districts after the seminar concludes. We'll

provide continued technical support and assistance for participants for the duration of the grant (9/30/2010) by continuing to support the four undergraduate students, who will maintain contact with the participants and build and develop the project's website. Along with Dr. Cream at the Center for Digital Humanities, the project faculty will make themselves available to continue the processes of mentoring and collaborating in which this seminar invests so heavily. Our project's web site, at <http://cdh.sc.edu/handheldart>, will continue to provide a mechanism to foster meaningful collaboration and technical development long after the seminar ends.

### **Project Faculty and Guest Lecturers**

Dr. Karen Heid (Project Director) is an assistant professor of art education at the University of South Carolina, working in secondary art education, multiage learning, and technologies of art. Dr. Heid has over 20 years of art making, art instruction, teaching, and leadership experience in K-12 schools and colleges in Georgia and South Carolina. Dr. Heid's research centers on multiage teaching and learning in the K-12 classroom, aesthetic development for students in the K-12 art classroom, and specifically postmodern curricular issues in developing arts unit of instruction. As an active and practicing scholar-artist, Dr. Heid has written curriculum for the High Museum of Art in Atlanta, The Performing Arts Workshop in San Francisco, Napa Valley Arts Council, and the Palo Alto Arts Center. Dr. Heid has shown her work in many different venues throughout the southeast and her work is in the permanent collections of many arts venues, homes and offices throughout the southeast. Dr. Heid has taught art classes in the both the private sector and in after school programs. She owns her own private studio and continues to teach private art lessons. These accomplishments on a personal level have prepared her for, and continue to inform her studies in higher education. The selected courses Dr. Heid teaches at the University of South Carolina include classes on art methods,

teaching, aesthetics, art criticism, service learning, and curriculum. With an emphasis on teaching issues related to contemporary curriculum, Dr. Heid works with students to develop curricula for the postmodern art classroom as the basis for a contemporary art classroom. Utilizing mobile technologies in the 9-12 art classroom will be important to developing a contemporary and postmodern art classroom curriculum. Dr. Heid will be providing the bridge between the mobile technologies and the existing curricular practices so as to create a new approach to new learning.

As a theory-oriented new media scholar in the Department of Art/Media Arts and Film and Media Studies Program at the University of South Carolina, Dr. Heidi Rae Cooley (co-director) is in a unique position to facilitate critical discussion regarding emerging mobile technologies and youth culture. Her research inquires into the mutually constitutive domains of poiesis (creative production), aesthesis (sensory knowing), and ethos (practice of living), particularly as these pertain to everyday practices with mobile screenic devices (e.g., camera phones). She has written extensively on the tactile connection between user and mobile device, underscoring the ephemeral nature of the products of mobile phone culture. She also holds a California single-subject teaching credential and, in addition to almost a decade of experience in secondary classrooms, Dr. Cooley brings her early work as a teacher trainer. Her multiple avenues of research and training make her ideal to help secondary art teachers reconsider the pedagogical uses of mobile technologies.

Dr. Randall Cream, co-director, is Associate Director of the Center for Digital Humanities at the University of South Carolina. With active projects involving educational gaming, mobile platforms in social networking, and the ethics of participatory learning, Dr. Cream is well-suited to serve as technical leader of the seminar. Trained extensively in

philosophy and cultural theory, Dr. Cream will also work closely with Dr. Cooley on developing a sophisticated sense of the theoretical impetus for handheld culture. Dr. Cream will direct the graduate assistant and the undergraduate students working hand-in-hand with the project participants, ensuring that the barriers to engaging in curricular change do not derive from the technologies involved. Dr. Cream will work closely with each of the project members to ensure the sustainability of the seminar's work going forward.

Simon Tarr, co-director, is a media practitioner and theorist in the Art Department's Media Arts and Studio Arts division at the University of South Carolina. Prof. Tarr will be providing crucial design support for the seminar participants as they work toward developing educational content for handheld devices. As an accomplished film maker, game designer, and imaging specialist, Tarr is well-versed in facilitating discussions of mobile movies, moving pictures, and image capturing for handheld devices. Moreover, given his theorizing of interactivity, he will invite participants to imagine participatory learning with handheld devices as a mode of a performative enactment of ideas. Prof. Tarr is a skilled technologist, as well, and will provide crucial support to the seminar participants.

#### Guest Lecturers

B. Stephen Carpenter II, PhD (Guest Lecturer) is associate professor of art education at Texas A & M University and a scholar-researcher in art education. Dr. Carpenter brings an expertise in curriculum development, curriculum theory, contemporary visual culture pedagogy, digital technologies, and interdisciplinary studies in art education. Teacher participants will benefit in multiple ways from Dr. Carpenter's leadership and teaching through his passion for the academic subject, his exemplary teaching skills, and his talent for communicating ideas. Dr. Carpenter has experience speaking nationally and internationally and has made numerous

presentations at professional meetings and conferences for art education and computer technologies. In addition to his professional research and publishing, he teaches courses at Texas A & M in creative inquiry and curriculum development. Dr. Carpenter also teaches courses in contemporary visual culture as an interdisciplinary investigation of related cultural, social, political, digital, ontological, and education issues, theories, and production and consumption practices in the postmodern era. Throughout his work, Dr. Carpenter examines contemporary visual culture as a site of critical inquiry that promotes social justice, cultural work, and democratic pedagogy.

Robert Quinn, PhD, (Guest Lecturer), is an assistant professor of art education at East Carolina State University. Dr. Quinn brings essential knowledge in mobile technologies and hands on practice for the 9-12 grade art classroom teacher. His selection as a guest lecturer for this seminar is based on his knowledge and expertise in art education areas of networking, vodcasting, podcasting, web design, and other computer technologies principally for art education practice. Dr. Quinn suggests that the most powerful way to strike a balance in large classrooms is the use of interactive computer technologies. Selected courses that Dr. Quinn teaches are computers in art education, art appreciation, methods of teaching, and curriculum development. In his computer courses students utilize computer technologies for the purposes of enhancing portfolios, preparing instructional tools, and developing curriculum in the digital arts. Dr. Quinn discusses pertinent digital technologies, visual culture, and contemporary classroom theories in art education and other related disciplines for K-12 instruction. His students work together in communities of practice through shared inquiry as inclusionary models of computer technology in the art classroom. Dr. Quinn will help the teacher participants move from

practicing the art of mobile technologies to writing short programs, creating lesson plans and empowering teacher participants to be ready to teach what they have learned to other faculty.

Pamela G. Taylor, PhD (Guest Lecturer) is Chair and associate professor of art education at Virginia Commonwealth University. Dr. Taylor, as an influential and precursor researcher of computer technologies in art education, has contributed to both instrumental and transformative to the vision and awareness of digital technologies in art education. In her research and teaching, Taylor suggests that our lives have been increasingly immersed in such technologies as television, film, and interactive technologies and we have found ourselves inundated with a heightened sense of mindfulness. Drawing on these ideas, Dr. Taylor believes that aesthetic experiences are made possible through such technological characteristics as hyperlinks, hypermedia, and hyper-reality. In this way, Taylor's work is at the forefront of art as participatory learning. We believe that Dr. Taylor would inherently benefit the teacher participants thorough her experience and knowledge of art educational practices in digital technologies with her knowledge and ability to connect teaching and learning to the digital and networked art education world.

### **Participant Selection**

Our seminar will work to draw a broad-array of art educators to form a community of engaged learners at the University of South Carolina. We'll advertise broadly, hoping to inspire art educators to initiate curricular change. Our selection process for participants uses local educators and project faculty to read through the application materials to build a diverse group who can work together and field an array of talented teachers. Applicants will write a narrative of their technological practices and investment in participatory learning, enabling us to select the most qualified and committed group of teachers for the seminar.

## **Professional Development**

With a team of faculty who are well-accustomed to the licensure processes of teaching, our seminar will be well-prepared to assist participants in receiving CEU credits for the work of the seminar.

## **Institutional Profile**

The University of South Carolina is fortunate to be able to offer the wide variety of faculty and staff needed to undertake this summer seminar. With an active and ambitious Center for Digital Humanities, an Art Education program interested in working at the leading edge of teacher preparation, and a Media Arts program that values both producing media art and understanding its meaning and value, our institution can supply the specialists needed to facilitate this seminar. During the three-week session, we will take advantage of the facilities of the department of Art, which has purpose-built classrooms designed to facilitate collaborative work such as ours. We are well prepared with the computing facilities to support fifteen participants during the three-week institute; we intend to take advantage of this redundancy to provide multiple overlapping venues for small group interaction, after hours sessions, and weekend development. The Center for Digital Humanities will provide the infrastructure to host the participants' work projects, through a shared server, allowing the participants to extend the work of the institute far beyond its temporal and spatial confines. We will happily support our core of fifteen participants, and just as gladly extend the same support to new partners as they emerge. Facilities are not the limiting feature at our institution to an endeavor such as this; we look forward to sharing in the rich intellectual milieu afforded by the institute. The university is able to make available conference-style housing for the participants in the student dormitories, and there are numerous hotels within a few blocks of the university.