



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES

Sample Application

Program: Digital Humanities Start-Up Grants

Note: The attached sample application was awarded a grant during a previous competition. Note that resumes, letters of support, coversheets, and other pieces of the application that contain personal contact information have been removed.

Significance and Impact

Today nearly *all* literature is *born digital* in the sense that it is composed with a word processor, saved on a hard drive (or other computational storage media), and accessed as part of a computer operating system. This new technological fact about writing is already having an impact, from office work to government and the academy to literature and the creative arts. In the particular realm of literature and literary scholarship, this means that a writer working today will not and *cannot* be studied in the future in the same way as writers of the past, since the basic material evidence of their authorial activity--manuscripts and drafts, working notes, correspondence, journals--is, like all textual production, increasingly migrating to the electronic realm.

What are the boundaries of authorship in an era of blogs, wikis, instant messaging, and email? Is an author's Web browser history part of her papers? What about a chat transcript or an instant message stored on a cell phone? What about a character or avatar the author has created for an online game? Should a researcher get to see an author's high score on Tetris or their choice of desktop wallpaper? Should a researcher get to see the songs available in an author's MP3 playlist? What of the author's choices for fonts and layout as expressed by the Preferences in their software? Such details may seem *recherché*, but in fact scholars regularly want to know what music an author was listening to or what images were in front of them during the writing process. The growing effectiveness of forensic information techniques raise the stakes still further: is erased or overwritten data appropriate to include within the scope of a literary collection if it can be recovered in a usable form?

Such questions have already passed from the realm of the speculative to ground-level decision-making about accessioning, curating, and preserving the born-digital literary records of today for future scholarly use. Digital Humanities Initiative Level 1 Start Up funding is therefore requested to support a series of site visits and planning meetings among personnel working with the born-digital components of three significant collections of literary material:

- The Salman Rushdie papers at Emory University's Woodruff Library--the 98-box collection includes a desktop (Mac Performa 5400/180), 3 laptops (a Mac PowerBook 5300 and 2 Mac PowerBook G3s), and a FireWire external hard drive, containing email, faxes, manuscript drafts, financial records, photographs, graphics (including at least one self-portrait), notes, and travel plans.
- The Michael Joyce Papers at the Harry Ransom Humanities Research Center at the University of Texas at Austin, which include 371 3.5-inch floppy disks, three CDs, two DVDs and one laptop, all containing correspondence, notes, journal entries, and manuscript material, including the particularly fascinating instance of a novel that is represented both by holograph and typescript as well as a variety of electronic formats ranging from MS Word to HTML to TK3 e-book.
- The Deena Larsen Collection at the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland, which includes eight Mac Classics, a Mac Plus, and associated peripherals and accessories, as well as some 800 diskettes, nearly 100 CD-ROMs, and several dozen Zip disks.

The meetings and site visits will facilitate the preparation of a larger collaborative grant proposal among the three institutions (likely submitted to NEH and/or IMLS) aimed at developing archival tools and best practices for preserving and curating the born-digital records of contemporary authorship. Initial findings will be made available through a jointly authored and publicly distributed white paper, as well as conference presentations at relevant venues.

Table of Contents

| | |
|--------------------------------------|----------|
| Statement of Significance and Impact | 1 |
| Table of Contents | 2 |
| List of Participants | 3 |
| Narrative | 4 |
| Biographies | 16 |
| Budget | Attached |
| Letters of Commitment | Attached |

List of Participants

Farr, Erika, Robert W. Woodruff Library, Emory University.

Kirschenbaum, Matthew G. (Principal Investigator), University of Maryland.

Kraus, Kari M., University of Maryland.

Nelson, Naomi L., Manuscript, Archives, and Rare Book Library (MARBL), Emory University.

Redwine, Gabriela, Harry Ransom Center, University of Texas at Austin.

Stollar Peters, Catherine, New York State Archives.

Approaches to Managing and Collecting Born-Digital Literary Materials for Scholarly Use

Digital Humanities Initiative Level 1 Start Up funding is requested to support a series of site visits and planning meetings among personnel working with the born-digital components of three significant collections of literary material: the Salman Rushdie papers at Emory University's Woodruff Library, the Michael Joyce Papers at the Harry Ransom Humanities Research Center at the University of Texas at Austin, and the Deena Larsen Collection at the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland. The meetings and site visits will facilitate the preparation of a larger collaborative grant proposal among the three institutions aimed at developing archival tools and best practices for preserving and curating the born-digital documents and records of contemporary authorship. Initial findings will be made available through a jointly authored and publicly distributed online white paper, as well as conference presentations at relevant venues.

Enhancing the Humanities

Today nearly all literature is *born-digital* in the sense that it is composed with a word processor, saved on a hard drive (or other computational storage media), and accessed as part of a computer operating system. Often the text is emailed to an editor, along with ancillary correspondence. Editors edit electronically, inserting suggestions and revisions and emailing the file back to the author to approve. Publishers use electronic typesetting and layout tools, and only at the very end of this process is the electronic text of the manuscript (by now the object of countless transmissions and transformations) produced as the static material artifact that is a printed book. This new technological fact about writing is already having an impact, from office work to government and the academy to literature and the creative arts. In the particular realm of literature and literary scholarship, this means that a writer working today will not and *cannot* be studied in the future in the same way as writers of the past, since the basic material evidence of their

authorial activity--manuscripts and drafts, working notes, correspondence, journals--is, like all textual production, increasingly migrating to the electronic realm.

What are the boundaries of authorship in an era of blogs, wikis, instant messaging, and email? Is an author's Web browser history part of her "papers?" What about a chat transcript or an instant message stored on a cell phone? What about a character or avatar the author has created for an online game? Should a researcher get to see an author's high score on Tetris or their choice of desktop wallpaper? Should a researcher get to see the music available in an author's MP3 playlist? What of the author's choices for fonts and layout as expressed by the Preferences in their software? Such details may seem mundane, but in fact scholars regularly want to know what an author was listening to or what images were in front of them during the writing process. The growing effectiveness of forensic information recovery raises the stakes still further: is erased or overwritten data appropriate to include within the scope of a literary collection if it can be recovered in a usable form?

Such questions have already passed from the realm of the speculative to ground-level decision-making about accession, curation, and preservation as writers have begun including diskettes, CD-ROMs, even entire computers and laptops in their literary papers. To date, however, the activity associated with processing such born-digital material has been localized and idiosyncratic, without much cross-communication among different archives and repositories; moreover, the archives and repositories, for their part, have not yet broached the kind of questions that are posed above with the scholars who will seek to access born-digital literary material in the years to come. Literary scholars are going to need to play a role in decisions about what kind of data survives and in what form, much as bibliographers and editors have long been advocates in traditional libraries settings, where they have opposed policies that tamper with bindings, dust jackets, and other important kinds of material evidence. This start-up grant therefore proposes to bring together library and academic personnel associated with three significant born-digital collections (described further in the next section) for a series of targeted site visits and planning

meetings at each of their respective institutions, with the goal of working towards a larger project proposal designed to address the needs of both archivists and scholars in this new literary milieu.

History, Scope, and Duration

The three participating collections are:

- The Salman Rushdie Papers at Emory University's Woodruff Library. The 98-box collection acquired in 2006 includes manuscripts, letters, notebooks, photographs, and, most significantly for our purposes, four computers with peripherals: a desktop (Mac Performa 5400/180), 3 laptops (a Mac PowerBook 5300 and 2 Mac PowerBook G3s), and a FireWire external hard drive. The desktop machine includes drafts and other material related to *Haroun and the Sea of Stories*; *East, West: Stories*; *The Moor's Last Sigh*; *The Ground Beneath Her Feet*; *Midnight's Children*; and *Step Across This Line*. The staff estimates that there are 690 MB of data stored on it (including everything). The external hard drive contains mostly snapshots of Rushdie's family and activities. All of the computers contain a rich variety of files and formats. The files include email, faxes, manuscript drafts, financial records, photographs, graphics (including at least one self-portrait), notes, and travel plans. In several cases, there are folders with back-up copies of files that allow us to see earlier versions of selected documents. Log files and software will help researchers complete a portrait of the way Rushdie used his computers for composition and revision, as well as other aspects of his writing life.
- The Michael Joyce Papers at the Harry Ransom Humanities Research Center at the University of Texas at Austin. The Michael Joyce Papers are composed of both digital and paper versions of published and unpublished works; academic,

administrative and teaching materials; correspondence; software documentation; journals and appointment books; student papers and personal documents. Highlights include Michael Joyce's published and unpublished fictional works, editor and agent correspondence and other materials related to his literary career. Major works represented include *Afternoon, a Story* (1987), a hypertext novel, and *The War Outside Ireland* (1982) and *Liam's Going* (2002), both linear print novels. Because *Afternoon, a Story* was composed almost entirely in the digital environment, very few paper draft materials exist within the collection. One of the most interesting portions of the collection is the creative progression of the novel *Going the Distance* (2002), from printed drafts (holograph and typescript), to a Microsoft Word version, to HTML, to a final version of the novel as published by NightKitchen in their TK3 reader software. The physical media on which electronic records within the collection were accessioned consist of 371 3.5-inch floppy disks, three CDs, two DVDs and one laptop of stored digital manuscript material that has been transferred from the original media to the Ransom Center's institutional electronic repository. Most of the electronic media was formatted for Macintosh operating systems, OS8, OS9 and OSX. Most of the hypertext works in the collection were created using specialized software, such as Storyspace. Versions of Storyspace software are available in the Storyspace section of the institutional repository (meaning that researchers must depend on objects stored in one portion of the electronic collection to access materials elsewhere in the same electronic collection).

- The Deena Larsen Collection at the Maryland Institute for Technology in the Humanities (MITH) at the University of Maryland. The Larsen collection, the 2007 gift of Deena

Larsen (an independent writer and scholar who has been active in the creative electronic writing community nearly since its inception) consists of an extremely diverse array of hardware, electronic data, and traditional literary materials, such as manuscripts, notes, correspondence, and ephemera. The collection includes not only Larsen's own extensive literary output, but original and sometimes unpublished material by nearly every significant member of her circle, effectively making it a cross-section of the electronic writing community during its key formative years (roughly 1985-1995). The files contain multiple versions of Larsen's *Marble Springs* and other Hypercard works (unpublished), multiple versions of her *Samplers* and other Storyspace works (unpublished), multiple versions of the poet William Dickey's works, Dickey's student works, poet Stephanie Strickland's works, M.D. Coverley's works, Kathryn Cramer's works, *If Monks had Macs*, the *Black Mark* (a hypercard stack developed at the 1993 ACM Hypertext conference), *Izme Pass*, Chris Willerton's works, Mikael And's works (the author himself no longer has copies), Jim Rosenburg's works, Michael Joyce and Carolyn Guyer's works that were in progress, Stuart Moulthrop's works, George Landow's works and working notes, textual games from Nick Monfort, *Coloring the Sky* (a collaborative work from Brown in 1992-94), and Tom Trelogan's logic game. The hardware in the collection consists of eight Mac Classics, a Mac Plus, and associated accessories; the physical media includes some 800 diskettes, as well as nearly 100 CD-ROMs and Zip disks.

While this is a new project, several of the participants have worked together before, and the project builds on converging interests among the members of the group. Kirschenbaum was the first scholar to use the Michael Joyce Papers at the Harry Ransom Center, and spent a week there working with Catherine Stollar Peters in February 2006. Kirschenbaum's initial findings from that visit are documented in his *Mechanisms: New Media and the Forensic Imagination* (January 2008, MIT Press). In addition, Kirschenbaum and Kraus are both co-participants in a Library of

Congress NDIIPP funded project on Preserving Virtual Worlds, which promises to engage a number of parallel issues. Kirschenbaum's "Hamlet.doc: Literature in a Digital Age" was the cover story in the August 17, 2007 issue of the *Chronicle Review*, and lays the conceptual groundwork for the activity described in this proposal.

Since the purpose of the start-up work is to create the foundation for a larger grant proposal, we expect and assume that this work will be the beginning of a long-term collaboration among the participants and their institutions.

The University of Maryland is the lead applicant and will administer the grant.

Staff

The proposal brings together personnel from two well-known institutional repositories as well as a working digital humanities center (MITH, which is also the home for one of the collections). Moreover, as the staff biographies reveal, a number of the participants hold advanced degrees in English or other humanities fields as well as library and information science; both of these factors will help ensure that the work transcends the relatively narrow concerns of library and information science.

Matthew Kirschenbaum, the Project Director, has several highly visible publications related to the central concerns of this proposal, including a cover story entitled "Hamlet.doc: Literature in a Digital Age" in the August 17, 2007 issue of the *Chronicle Review*. He has long-standing interests in digital preservation, which he has written about in his forthcoming *Mechanisms: New Media and the Forensic Imagination* from the MIT Press (early 2008) and he is the individual chiefly responsible for the acquisition of the Larsen Collection at MITH. His interests and expertise are well-complemented by the other project members, who collectively bring credentials in literary studies, information science, copyright and intellectual property, digital preservation, archival science, and textual editing.

At both the Ransom Center and Emory's Woodruff Library, the individuals who are most directly charged with working with the born-digital collections are those who will be participating in the grant. As their vitas indicate, Erika Farr, Naomi Nelson, Catherine Stollar Peters, and Gabriela Redwine are all active information professionals presenting papers and running workshops at major conferences on topics directly relevant to the work proposed here (Redwine has experience with Norman Mailer's electronic records, as well as with the Joyce Papers). And although she is no longer directly affiliated with the Ransom Center, Stollar Peters's participation is especially valuable since she was the individual responsible for most of the initial processing of the Joyce collection; she maintains a strong interest in the material. Kari Kraus combines expertise in textual scholarship and information science with an understanding of intellectual property that will be essential. In sum, we believe we have created a very strong interdisciplinary team that is uniquely situated and qualified to address the challenges presented by this material.

Methods

The goal of the site visits and meetings will be to familiarize the principals with one another's collections in their local context, and to generate consensus for a larger grant application that will fund applied work in developing tools, best practices, and strategies for librarians, archivists, and scholars working with born-digital literary collections in the future. Each meeting will include an extensive hands-on orientation to the born-digital collection at the host institution, allowing members of the group to experience the environment in which the collection will be used and to discover idiosyncrasies of the materials for themselves. Moreover, we will take advantage of the setting of each of the three meetings in College Park, Austin, and Atlanta to meet with a diverse group of local experts. While specifics are necessarily tentative at this stage, the following are three prospective agendas, starting with the first meeting, which would be held at the University of Maryland in September of 2008. (The agendas are included here rather than as appendices because they offer the most concrete statement of our methods.)

- 8:30 – 9:00. Introductions, review of objectives for grant.
- 9:00 – 11:00. Overview of the Larsen Collection, including scope and content of the collection, literary significance, summary of work completed to date, planned future workflows, problems and challenges identified, and use cases.
- 11:00 - 12:00. Briefer overviews of the other two collections (when on site at the other two locales, the Joyce and Rushdie collections will each receive an overview as extensive as the Larsen materials here).
- 12:00 – 1:00. Working lunch with members of the University of Maryland creative writing faculty, such as Michael Collier (Professor of English and Former Poet Laureate of the State of Maryland) and Maud Casey (Assistant Professor of English and author of two published novels and a short story collection). The goal will be to interview them about their electronic composition practices.
- 1:00 – 2:00. Meeting with a member of the Library of Congress's Section 108 Study Group to discuss copyright and intellectual property as they relate to born-digital archiving and preservation.
- 2:00 – 3:00. Meeting with Bruce Ambacher, Professor, College of Information Studies (electronic records specialist) and his graduate students.
- 3:00 – 4:30. Identification of emerging issues, discussion of next steps and agenda items for the Austin meeting, to follow in 4-6 weeks.

Likewise, a prospective agenda for the second meeting, at the Harry Ransom Center in Austin in middle or late October 2008:

- 8:30-9:30. General overview of collections at the HRC that have electronic records and the work that has been completed on them to date.
- 9:30-11:00. Overview of Michael Joyce Collection, including scope and content, significance (literary and archival), variety of tools used to access and preserve

information in the collection, work completed to date, major challenges and solutions.

- 11:00-12:30. Demonstration of Michael Joyce material on DSpace, analysis of DSpace successes and challenges, opportunity for participants to access materials directly using computers in the reading room.
- 12:30-1:30. Lunch.
- 1:30-2:30. Discussion of use, overview of HRC's experiences with providing access, incorporate participants' experience with Joyce materials.
- 2:30-3:30. Meeting with Pat Galloway, Associate Professor (and digital preservation/electronic records specialist), School of Information, and Ph.D. students conducting research projects on digital preservation.
- 3:30-5:00. Conclusions, questions, next steps.

Finally, a prospective agenda for the third meeting, at Emory's Woodruff Library in Atlanta in early December 2008:

- 9:00-9:15. Introduce any Emory colleagues who have not been at other meetings (such as Library Director). Review agenda.
- 9:15-10:00. Updates and review findings from previous meeting.
- 10:00-12:00. Overview of Salman Rushdie's Collection. Description of the traditional and digital components of the archives and our processing of these collections. Discuss the questions and challenges Emory has faced with the handling of this archive. Focus on issues of accessibility and how scholars might engage such a rich collection.
- 12:00-1:00. Lunch.
- 1:00-3:00. Comparative Review of how participants have processed and provided access to born-digital material. After learning about born-digital archives at all three institutions,

the group can discuss differences in approach and result. Through comparison, explore emerging best practices and trends.

- 3:00-5:00. Planning subsequent collaborative initiative around born-digital archives and scholarly use. Discuss strategies that will capitalize on the planning and conversations of three meetings. Begin planning to develop proposal for collaborative project.

Each meeting will include extensive notes and record-keeping. It also goes without saying that the visits will allow for considerable informal contact among the participants, allowing conversations to continue to develop away from the meeting table. Each day-long meeting will be followed by a dinner with all of the project members.

As vexed as the issues surrounding the born-digital materials are, in practice the situation is even more complex, since in all three collections the electronic data is part of a hybrid acquisition that includes print and other physical media manifesting important and subtle relations to the digital files. The Michael Joyce Papers, for example, include the manuscript for a work entitled *Going the Distance* which exists as a holograph and a typescript as well as in a variety of electronic formats including MS Word, HTML, Storyspace, and as a proprietary eBook format. A scholar studying the composition of such a text will require access to all of its versions, as well as the tools to navigate effectively between them. The Rushdie Papers similarly yields manuscript material for important works in both physical and virtual form. The Larsen collection includes material as heterogeneous as a shower curtain to which laminated “screens” of the Hypercard *Marble Springs* are affixed and connected by pipe cleaners to tangibly map the work’s system of links and relations. We therefore expect an emphasis on the hybridity of these materials (physical and virtual records commingling and mutually informing one another) to be a signature aspect of our long-term efforts.

Final Product and Dissemination

The most immediate product of the meetings and site visits will be a lessons-learned white paper jointly authored by the participants, addressing at least the following topics: authorial practices in the digital age; scholarly needs; preservation challenges; copyright and intellectual property; privacy; forensic data recovery; born-digital content as a “special collection”; network access; security; archiving hardware; hybrid print-digital collections. The white paper will be publicly available online, as described in the NEH guidelines for the Start-Up award.

Beyond that, our hope and expectation is that the white paper will furnish the basis for a much larger grant proposal, submitted in the second half of 2009 to the NEH, the IMLS, or the joint IMLS/NEH Archiving Knowledge program. Our working assumption is that such a grant would fund the building of an archivist’s toolkit—which would include, for example, utilities to track the relations between the versions and components of a complex hybrid work like Joyce’s aforementioned *Going the Distance*—as well as guidelines for best practices in curating and preserving born-digital documents and records of literary and cultural significance for scholarly use.

We further assume that participants will derive their own individual papers based on issues raised during the meetings, and present these at appropriate professional venues including the Modern Language Association, the American Library Association, and the Society of American Archivists (to name a few). Given the widespread interest in the topic of preserving born-digital documents and records of literary and cultural significance, we expect these activities to be quite visible.

Work Plan

The meetings will be conducted in fall of 2008, approximately four to six weeks apart. The first is tentatively schedule for September 18-19, 2008 at the University of Maryland in College Park. The Austin meeting will follow in middle or late October, and the final meeting will be held at Emory University in Atlanta by early December. Participants will then jointly draft the white

paper and this will be available no later than March 1, 2009. All of the participants have agreed to this schedule.

We believe that before very long the three collections described here will no longer be regarded as special cases, but will in fact seem merely representative as more and more contemporary writers of literary significance seek to consign “papers” that also consist of personal computers, hard drives, diskettes, and other data storage media to libraries and archives. The repositories, for their part, will increasingly be faced with the question of how to curate and preserve this material in a manner that will be rich and meaningful for the scholars and students interested in an author’s career. We look forward to beginning the process of examining how to best meet those challenges, balancing the needs of scholarship and preservation a new, hybrid textual and curatorial environment.