

## 4. Narrative

### Enhancing the humanities through innovation.

Our Level II National Endowment for the Humanities (NEH) Digital Humanities Start Up grant proposal emerges from the Consortium for Electronic Literature (CELL), a partnership founded by the Electronic Literature Organization (ELO) and joining nine research centers, all developing online database projects devoted to research in electronic literature (e-lit). The ELO is the leading organization devoted to facilitating and promoting the writing, publishing, and reading of literature in electronic media. The ELO defines e-lit as “born-digital works with recognized literary aspects that take advantage of the capabilities and contexts provided by the stand-alone or networked computer.” These databases are the primary bibliographic resources available to scholars and students in the emerging research field of e-lit and represent diverse linguistic, cultural, and aesthetic approaches. Our international team includes leading e-lit scholars, as well as experts in database design, information/library science, and digital humanities project management.

Building on existing CELL achievements, we request **\$59,973** to implement the following:

1. **a web-based search engine for e-lit** to display aggregated server results harvested from across the consortium databases;
2. **a unified name authority system** to improve the data harvested by the search engine and to create more faceted and complex results;
3. **and a training/how-to framework** to extend our initiative to include future projects and partners and to establish standards and best practices in using the e-lit search engine.

All CELL projects use open-source tools and are committed to open access. Grant activities will last for eighteen months, from May 2014 through November 2015.

As a Level II proposal, our project is well underway. During the 2012-13 academic year, the consortium established a metadata standard for creative works of e-lit, the first of its kind. All CELL members signed agreements to configure their databases and to make records available in accordance with the standard. The consortium subsequently agreed to create a shared server to harvest records from across the CELL partner databases. The server is already in place and will use the open source Apache SOLR framework to index records. The individual partners are completing the configuration of their databases. The server will begin harvesting records by May 2014. Support from the NEH will ensure that this major step is maintained throughout the grant period. More importantly, NEH support will enable us to make significant innovations in e-lit through research with the server.

Our project’s objectives will allow scholars from within the field and in the general public to search for and retrieve richly connected research on e-lit. While stand-alone databases exist—*indeed, our consortium represents all the significant projects*—no consistent standard exists for the terms and notations used for encoding and retrieving the research in these databases. Some of the partners used highly controlled data, while others use more open, folksonomic approaches. Some of the partners are focused on very specific corpora of e-lit, while others attempt to encompass the entire field. By standardizing, aggregating, and making searchable data across the partner projects, we will allow researchers in e-lit—present and future—to build on, implement, and extend our inter-site searching. Moreover, our goal is to go beyond merely creating an “e-lit Google.” We will provide the first shared tool for curated, international research in the e-lit field. Our site will provide the go-to point for new scholars, as well as students and the general public. The search engine will display results from the aggregated data on the server, with direct links from items in the search to the full data in the individual projects.

Designing the search interface is not a neutral task but one that will require careful critical and scholarly attention. For example, the shared server will foreground the multilinguistic challenges of born digital writing in global collaborations. A guiding question will be how to represent and handle the crosslinguistic search results. Our decisions in designing the search interface will allow scholars to pose research questions at the intersection of ubiquitous and networked born digital objects, on the one hand, and the specificity of individual languages and cultures, on the other.

A related and equally important research question will be the use of taxonomies. In planning the shared server, we have required each partner database to provide their current taxonomies. To make the search process meaningful and scaleable as we add more partners, we will need to determine taxonomies to be implemented on the shared server. We will focus on this task once the shared server is in place, using grant funds to enable discussion, development, and testing of shared taxonomies. In turn, these taxonomies—stating, for example, the list of media in which e-lit works are published—will provide standardized lists for describing works in the field.

No name authority system exists for the e-lit field, a state of affairs that causes confusion and redundancies in the scholarship. While e-lit author names obviously intersect with existing name authority projects, such as VIAF, there is a clear need for a name authority system unique to e-lit. The emerging field uses terms and names in a vague and contested manner, with no clear agreement. For example, Marjorie C. Luesebrink is the author of many well-known works of e-lit. Some of these are published under her own name, but many are published under the pen name M. D. Coverley. As a result, e-lit databases include entries for both names. The matter is even more complex: there may be a separate entry for Marjorie Luesebrink (without the middle initial) or for MD Coverley (without the periods and spaces around the initials), often as a result of different catalogers. This leads to considerable confusion, since there is no way to gather and survey all the works by this author. A name authority system provides a thesaurus, mapping all instances of authors' names and allows catalogers to ensure that a reliable and agreed-upon name is used. It also allows the search engine server to provide a complex and faceted search, taking users from records listing M. D. Coverley as author to other records listing Marjorie C. Luesebrink as author. Our project will address these gaps and limitations by creating a shared interpretive framework that will be vital to the proliferation of research in this field. Scholars and students of e-lit will be able ask previously-unanswerable research questions about history, context, and reception.

Our goal is to make the server, search engine, and metadata feeds available to the general public. To achieve this, our design and testing includes representatives of the scholarly community and the broader public, along with the project team. We will use a broad usability test base to meet the needs of audiences ranging from veterans of the scholarship to undergraduates newly exposed to e-lit. We will also create a framework for introducing future projects to our initiative and enabling them to participate.

As a result, our project's innovation extends beyond the field. E-lit is a barometer of the trajectory of texts in digital environments and deals with concerns shared by virtually all digital humanities projects. E-lit works are "born digital" in part or in whole and pose significant challenges for metadata and archiving. The innovations in handling these complex artifacts—not only in a single database but across an international consortium of databases—will provide lessons and insights for a wide range of digital humanities projects. Our search engine will codify the overall trajectory of literature in digital environments, providing concepts, relationships, and organization for the emerging condition of texts today. As the type and number of e-lit works expands, practices of naming and organizing become a necessary and important scholarly exercise. Keywords and relationships are needed to locate boundaries within the field (such as boundaries between textual, visual, sonic, and even tactile media). In some instances, literary terms from the print archive can be brought forward; often, as new terms arise, they need to be recognized and critically discussed by scholars and creators, both of which are represented in the CELL membership. (See discussion in Environmental Scan below.)

## **Environmental Scan**

The major scholarly publications in the field, such as N. Katherine Hayles' *Electronic Literature: New Horizons for the Literary* (Notre Dame 2008) and Christopher T. Funkhouser's *Prehistoric Digital Poetry*

(Alabama 2007) deal with authors, works, and thematics. They do not deal with the metadata and information architecture involved in e-lit projects, although the scholarship informs such projects. Metadata and similar standards are a primary means of developing robust tools for e-lit research. Looking beyond scholarly publications and toward existing research tools, one finds advanced consideration of digital records and archiving, as one would expect of e-lit. The ELO was a pioneer in discussion of born digital records through initiatives such as "Acid Free Bits" and "Preservation and Access." Projects developed by the ELO community—such as "The Agrippa Files," developed at UC Santa Barbara by then-president of the ELO Alan Liu and former ELO board member Matthew Kirschenbaum—are among the paradigmatic digital humanities explorations of complex textual artifacts.

A number of significant projects explore the challenges of archiving born-digital records in ways that communicate with and parallel work on e-lit. One project we see as instructional is UC Berkeley's "Archiving the Avant Garde: Documenting and Preserving Variable Media Art," which developed strategies to handle the variability of Internet and digital art. Another similar project is the multi-institution "Preserving Virtual Worlds," an ongoing collaboration between a number of universities, now in its second phase, to develop standards for metadata and content representation to preserve virtual worlds and interactive fiction. We expect to benefit from both projects' approaches and lessons learned. For example, one outcome of "Archiving the Avant Garde" was a "formal notation system for scoring works of digital and variable media art," which provides a comparative model for our controlled vocabulary and metadata standard (as with our project, the notation system uses Open Archives Initiative - Protocol for Metadata Harvesting/OAI-PMH standards).

The name authority for e-lit—one of the major deliverables of our project—looks toward established projects such as Virtual International Authority File (henceforth VIAF). VIAF is an international service "providing convenient access to the world's name authority files" (see <http://viaf.org>). Name authorities organize, match, and merge names and other bibliographic data to provide a single reference source. VIAF provides a "super" record, allowing reference across multiple organizations. Our project will create a similar authority across the CELL e-lit databases.

The e-lit field is developed enough to support a number of international research teams working with databases. These teams form the consortium at the center of our project. CELL is an initiative of the ELO to develop partnerships among organizations, universities, and publishers to share research on e-lit. CELL members represent the major players in the field of scholarly database-driven research into e-lit. The CELL agreement sets unified goals for the major international e-lit research teams, including: 1) to foster critical practice around e-lit through scholarly materials for research and reference; and 2) to advocate for the importance of online e-lit resources for scholarly research, including the development of such resources as a legitimate component of academic credentialing.

The following teams, all with research projects involving databases of electronic literature, have committed to the consortium and its objectives with signed letters of agreement:

- The ELO's Electronic Literature Directory (ELD);
- *The electronic book review* (ebr);
- Digital Arts Collection, Brown University Digital Repository;
- Creative Nation: Writers and Writing in the New Media Culture, University of Western Sydney (Australia);
- The Canadian Directory of Electronic Literature/Laboratoire NT2, Université du Québec à Montréal (Canada);
- The ELMCIP Electronic Literature Knowledge Base, University of Bergen (Norway);
- Media Upheavals / Literature, Art, and Media, University of Siegen (Germany);
- PO.EX - Digital Archive of Portuguese Experimental Poetry, University Fernando Pessoa (Portugal);
- Hermeneia, the Literary Studies and Digital Technologies Research Group, Universitat de Barcelona (Spain).

An agreement is in progress with the Laboratoire Paragraphe at the University of Paris VIII. A signed agreement is also in place with Archiveit.org, and through this the US Library of Congress for an ongoing

initiative to archive electronic literature in the LOC. This collaboration is a major indicator of the leading role played by the ELO in the institutional reception and understanding of born-digital literary works.

These individual projects represent multiple languages and national traditions. Moreover, not all are exclusively concerned with e-lit. For example, PO.EX concerns post-1960 Portuguese experimental poetry, some but not all of which is e-lit. Our project is not simply based in the narrow context of American academia. It has a rich historical and critical context where e-lit is represented along with works that relate to but are not, strictly speaking, e-lit. The consortium brings these disparate projects together for the first time. Ours is the first true cross-disciplinary attempt at formalizing the metadata and database practices in the field.

## History and Duration of the Project

The project builds on a previous NEH Level II Digital Humanities Start-Up grant to the ELO's Electronic Literature Directory, completed at the end of 2011, with the goal of improving and expanding the directory along semantic web models (see <http://directory.eliterature.org>). The project team reported a number of discoveries, including: deeper understanding of the tools and platforms used in e-lit; insights into the use of taxonomies representing the field; and insights into fostering critical analysis skills through educational applications of the ELD. The preliminary data from that project provides a starting point for our own project. However, the current project is a radically different one. It builds on the intellectual framework of the previous but employs a different team and is much larger in scope. For the first time, this project brings together the significant international research projects on e-lit. We are excited by the potential for new humanities discoveries through this initiative. The ELO also recently received an NEH grant for studying early works of e-lit: the "Pathfinders" project, directed by ELO Board Members Dene Grigar and Stuart Moulthrop, which will build an archive of readings in which authors and volunteer readers explore the textual possibilities of early digital texts.

The current initiative emerges directly from the CELL agreement, which emerged out of meetings organized by Anna Gibbs and Maria Angel of the University of Western Sydney in 2010, and a follow-up meeting at the University of Bergen in June 2011. Lisa Swanstrom of Florida Atlantic University coordinated the agreement between consortium members. Each of the e-lit projects in the consortium already has a large database, an established team, and a robust, multi-year initiative. The ongoing work on the project is coordinated in monthly partner meetings via Skype, with sub-committees meeting periodically to discuss topics such as taxonomies and authority files. Representatives of the consortium will meet in person at the ELO conference in Paris, September 22-27, 2013, to discuss current achievements and to plan forward progress.

A major achievement in the last year was "The CELL Metadata Element Set," approved by all consortium members in February 2013. (See the Appendices for a copy of the document.) It provides vocabulary terms for a standard representation of creative works of e-lit. The document emerged from discussion of the existing standards of individual member projects, with the intention of identifying a canonical minimum metadata description for a creative work of e-lit. Following this agreement, the consortium agreed to two targets for the time period up to the end of April 2014, as described below.

1. All partners will modify their databases to conform to make standard-compliant metadata available for harvesting. By the end of 2013, each partner will complete the process of providing metadata for harvesting.
2. The partners will create a centralized server using Apache SOLR to harvest records from the databases. Other solutions were discussed as alternatives to the centralized server, including representing the aggregated data on the individual partner sites. The decision was that while the latter alternative might prove more flexible in the long run, the first was desirable as a realistic goal, one that would provide the building blocks for subsequent developments (such as those set out in the current proposal). All the databases currently run on the open source Drupal content management system, the DSpace open source repository software, or provide access to their records via web API. There is an existing Apache Solr Search Integration Module for Drupal and DSpace has built-in Apache Solr integration. The goal is for the databases to be application-independent, both for long-term preservation and for compatibility with new consortium partners. The harvesting would follow

OAI-PMH. The NT2 Lab at the Université du Québec à Montréal, representing the Canadian Directory of Electronic Literatures, has provided a server for the project and is in the process of configuring it to aggregate the data. By May 2014, the server will be regularly harvesting across the partner sites.

## Work Plan

Our Level II project is well underway, as noted above. By May 2014, partner database projects will conform to the project metadata element set; the records will be available for harvesting; and the server at NT2 will be aggregating this data into a single index of the sites. In addition, all the partners will have met in September 2013 to sign agreements on supporting the server and committing to the project.

We have set the following benchmark dates for the next steps of the project.

### May-September 2014

- Continue to build the index. Each project will access the data and develop proposals for how to represent it on the web search engine. Monthly meetings will discuss the proposals.
- Create a minimal website as a placeholder until the group meeting in October (see below), displaying information about the server index. Grant funds will pay a stipend for a search engine website project leader, concerned with this task and others related to the search engine throughout the project (Gabriel Gaudette of the Université du Québec à Montréal).
- Begin to analyze name authority listings for all the partners. The server will create this aggregated file. We will then begin the process of comparing, collating, and mapping names. To achieve, we will use both manual and automated approaches (e.g. Google Refine/Open Refine). We will begin a similar process with the taxonomies provided by the partners. Grant funds will pay a stipend for a name authority project leader, concerned with this task and others related to the name authority throughout the project. (Peter Sefton of the University of Western Sydney)

### October 2014 (Four-day meeting at WVU)

- The meeting will include representatives from all the partners. In addition, subject matter experts (SMEs) —Professor Christopher Funkhouser of NJIT and Professor Rita Raley of UCSB, both leading scholars of e-lit—will be included in the meeting, as will Wendy Burk, Poetry Center Librarian from the University of Arizona. They will provide insight into the best ways to design and present the search materials, and will consult on the development of the name authority and taxonomies. Grant funds will provide lodging, per diem, and travel assistance for all meeting participants. *The expectation is that some of the travel costs will be paid by participants' home institutions.*
- The first two days of the meeting will focus on designing a web-based search engine using the server information, to allow users to search across all databases and consult the search results via direct links to the individual projects. All the partner teams will bring design documents indicating their specifications and needs for this first-ever e-lit search engine. We recognize that users need to see the provenance of individual items in a search and to understand the search as an integrated project reflecting a research consortium. At the same time, the individual projects will bring a variety of design needs based on their research priorities. These sessions will consist of rapid prototyping to complete an agreed-upon design of the website by the end of the meeting. We will leave the meeting ready to build the website/search engine.
- The second two days of the meeting will focus on the shared name authority system for the project and on the question of taxonomies. All existing names and taxonomies will already be collected from CELL members through the server. The goal of the meeting is to review the progress and resolve problems with the unified name authority system. A second goal is to arrive at agreement on taxonomies used in the server. We will also plan implementation of the name authority and taxonomies on the server and back to the partner sites. Additionally, we will discuss future coordination of the name authority with other authorities (e.g. VIAF). This task requires deep knowledge of the field and will necessarily draw on the expertise of the scholars on the partner projects as well as the consultant SMEs.

### November 2014-October 2015

- Create a framework allowing future projects to partner with and integrate into the consortium. We assume that more and more e-lit research projects will emerge—in fact, we know of developing databases in Russia and South America—and we must ensure that those new projects can easily adapt and be compatible with the consortium. Toward this end, we will create documentation, technical specifications, training materials, and video workshops directed to new researchers and research projects in the field. The materials will be made freely available via the Internet.
- The framework will allow new projects to answer basic questions about our initiative: How did we create the initiative? Does your project fit? How can you make your project a part of the initiative? (In particular: What do you need to do to your database to make it interoperable?) Grant funds will pay a stipend for a training/extensibility project leader, concerned with all aspects of the training framework throughout the project (Ewan Branda of Woodbury University and the ELD).
- *Ongoing through the end of the project:* we will complete and release the web search engine, and we will implement the name authority system, with support from the CELL partners.

### **May 2015-October 2015**

- Author and release project white paper. The white paper will include all materials from the project. It will be co-authored by representatives from all the teams, as well as the consultant SMEs.

In the CELL initiative's longer term development, beyond the current proposal, we intend to add feeds and services to the search engine so that individual projects can embed customized versions of the search in their site. We will also include material in the training framework on the long-term maintenance of the feeds from the individual projects to the server. A longer term goal not addressed in this proposal will be coordinating our name authority initiatives with other name authorities.

### **Staff**

The Project Director is Charles A. Baldwin (Sandy), director of the Center for Literary Computing at West Virginia University. He will be assisted by two student assistants, one paid by grant funds and one supported by WVU. Baldwin will be paid a stipend for his efforts during the summer of 2014 and 2015 (his WVU salary will cover his efforts during the academic year). Baldwin will devote 5% of his overall effort to the project. The WVU team will coordinate the project during the year: facilitate meetings and communication; maintain a web archive of discussions; and organize the October 2014 meeting.

The name authority task will be coordinated by Peter Sefton, the University of Western Sydney (Australia). The framework for extending the project will be coordinated by Ewan Branda of Woodbury University. The usability and design of the search engine will be coordinated by Gabriel Gaudette of the Université du Québec à Montréal. Each task leader will be paid a stipend for their efforts during the summer of 2014 and 2015 (we assume their home institution will fund their efforts during the academic year).

The project will draw on experts from outside the project team to provide additional oversight and consultation. Wendy Burk (University of Arizona) will provide subject-matter expertise from Library Science, while Christopher T. Funkhouser (New Jersey Institute of Technology) and Rita Raley (University of California in Santa Barbara) will provide subject-matter expertise from the e-lit field. All will attend the project meeting in October 2014.

Consortium teams will designate participants responsible for working group meetings and for communicating with individual project teams. The working group will draw on this broader expertise as needed. All participants will attend the October 2014 meeting in Morgantown, WV.

### **Final product and dissemination**

Final products include 1) the publicly-released search engine with validation/name authority; 2) the framework for extending the project; and 3) the final white paper. The white paper will be circulated to major online venues in the field, including DHQ, HASTAC, and EBR. Versions of the white paper will be presented at the 2015 ELO conference, to be held at the University of Bergen in Norway and at the Modern Language Association conference in January 2016.