



NATIONAL ENDOWMENT FOR THE

Humanities

OFFICE OF DIGITAL HUMANITIES

## **Narrative Section of a Successful Application**

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Office of Digital Humanities application guidelines at <http://www.neh.gov/grants/guidelines/IATDH.html> for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

**Project Title: Advanced Topics in TEI Encoding**

**Institution: Brown University**

**Project Director: Julia Flanders**

**Grant Program: Institutes for Advanced Topics in the Digital Humanities**

## Advanced Topics in TEI Encoding

### Significance

Among the many digital technologies that offer significant promise for advancing humanities scholarship, text markup is probably the most far-reaching and broadly useful. By “text markup” or “text encoding” we refer to a very broad range of representational mechanisms that serve digital scholarship, including various kinds of metadata, finding aids, structured image keywording, and digital name and subject authority records. Text encoding, broadly speaking, is thus an extremely powerful and flexible approach to representing many different kinds of important humanities data. However, the text encoding language with the greatest direct relevance to humanities scholars is the TEI Guidelines,<sup>1</sup> which is increasingly becoming an essential competence for scholars who wish to create high-quality digital humanities resources, editions, and other forms of digital scholarship. The challenge is in finding a way to learn it.

This challenge is gradually being met. Introductory text encoding workshops for humanities scholars are becoming more common; in North America there are a few regular annual events and an increasing number of occasional events<sup>2</sup> through which scholars and librarians can gain an initial acquaintance with the concepts and techniques of text encoding and the TEI. Over the past six years, the Brown University Women Writers Project staff alone have offered over twenty workshops and seminars, with a total attendance of approximately 300 humanities scholars, students, and librarians. Many of these people have gone on to create or participate in new digital humanities projects. However, they (and others with a basic knowledge of text encoding) face a further challenge in finding sources of advanced encoding instruction and guidance, focused more closely on the areas of the TEI which are specifically relevant to their projects. What is needed, for this cohort of digitally aware scholars and practitioners, is an opportunity to explore specific areas of TEI markup in more detail and at a more advanced level, in a context where they can also learn strategies of decision-making and project design. In addition, the most recent release of the TEI Guidelines contains substantial new provisions which are of particular relevance to this audience, but which are complex and not easy to digest in isolation.

Responding to this need, the Women Writers Project proposes a three-year series of advanced text encoding institutes, aimed primarily at scholars, librarians, and archivists who are currently engaged in a digital text project and have a basic working knowledge

---

<sup>1</sup> The Text Encoding Initiative Guidelines for Electronic Text Encoding and Interchange, published by the TEI Consortium, <http://www.tei-c.org>. Over the past two decades these guidelines have emerged as a crucial community standard for the digital humanities.

<sup>2</sup> Annual workshops include the Digital Humanities Summer Institute at the University of Victoria (<http://www.dhsi.org/>), and an annual workshop at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. Workshops have also been held at conferences such as the International Congress on Medieval Studies at Kalamazoo, or as special events sponsored by organizations like NITLE (see e.g. <http://www.wheatoncollege.edu/KACC/tei/jan05/projdesc.html>). The Women Writers Project offers regular workshops and seminars, including an NEH funded seminar series now in its second year; see <http://www.wwp.brown.edu/encoding/seminars/> and <http://www.wwp.brown.edu/encoding/workshops/>.

of the TEI. We will include the option of a one-day pre-institute intensive TEI introduction to enable participants who need a refresher to get up to speed, but the assumption will be that participants are already engaged in TEI encoding and are able to engage with more advanced questions and problems. The main body of each institute will be three days long; this length is sufficient to permit substantive progress and project development, without risking burnout or posing logistical difficulties for participants with other commitments. These institutes will build on the WWP's current NEH-funded seminar series (<http://www.wwp.brown.edu/encoding/seminars/>) and also on the other regular workshops we offer. They will also complement other introductory TEI workshops and courses being offered at library schools, conferences, and digital humanities programs. To our knowledge, there are no equivalent events available anywhere in North America.

## Institutional profiles

The Brown University Women Writers Project, which will provide the primary instructors for these institutes, is a long-standing leader in the field of humanities text encoding and especially in providing outreach, documentation, and support for new text encoding projects. We have received generous and repeated funding from the NEH for documentation (including our *Guide to Scholarly Text Encoding*, <http://www.wwp.brown.edu/encoding/guide/>) and training seminars, and we are well known as an authoritative and active source of guidance and support for individuals and projects. Our staff have unrivalled expertise in the TEI based on sustained engagement with the technical, theoretical, and organizational aspects of the TEI almost since its inception.

In ensure a broad geographical range, we are conducting the institutes in partnership with a set of institutional hosts located on the east and west coasts and in the center of the country. We have firm hosting commitments from institutions on the east and west coasts and from two institutions in the midwest and south. We have also had initial discussions with the University of Illinois at Urbana-Champaign in which they expressed interest in hosting, but time did not permit documentation of a full commitment by the proposal deadline. If this proposal is funded, an additional host in the center of the country will be recruited, probably UIUC, and an additional west-coast host as well. All of the host institutions have strong records of involvement in the digital humanities, and several have already served (or will serve) as hosts for the WWP's current TEI seminar series. The hosts are:

- **The University of California, Santa Barbara** currently has a number of important digital humanities projects under way, including the Transliterations Project and the English Broadside Ballads Project. Alan Liu and Patricia Fumerton provide strong leadership for UCSB's digital initiatives and they hosted a very successful TEI seminar in September 2008. Their facilities include two computer labs and seminar space in close proximity, as well as excellent technical support. They have agreed to host one institute.
- **The University of Kentucky** is an important center for research on advanced digital methods of working with manuscripts and ancient texts. The Collaboratory for Research in Computing for Humanities

(<http://www.rch.uky.edu/>) supports a number of important digital projects including EDUCE (a project focused on methods of virtual unwrapping and reading of fragile documents), Electronic Beowulf, Electronic Aelfric, and several others. RCH is willing to host an institute and would be an excellent venue. Dot Porter, the RCH Program Coordinator, has served on the TEI Council and on the TEI's Manuscript Description workgroup, and has also taught workshops on manuscript encoding. She would serve as local coordinator at RCH and also as a guest instructor for the manuscript-centered institutes.

- The **University of Maryland** has a long history of leadership in digital humanities, both through its digital library programs and also through the Maryland Institute for Technology in the Humanities (MITH), which has sponsored a number of very significant digital projects and tools, including the Dickinson Electronic Archive, the Versioning Machine, and the Virtual Lightbox. Its Director, Neil Fraistat, and its Associate Director, Matthew Kirschenbaum, are both distinguished scholars of the digital humanities, and have provided strong support for the WWP's current TEI seminar series. The MITH facility is well equipped for hands-on digital events such as this institute. They have agreed to host a series of three institutes.
- The **University of Nebraska-Lincoln** has emerged in the past five years as a strong locus of digital humanities projects and research, with major TEI projects such as the Walt Whitman Archive and the Willa Cather Archive. Its Center for Digital Research in the Humanities serves as the organizational core of this work and has access to good facilities for hosting the institute, including a good lab space with computers and a seminar space. Brett Barney, Project Manager for the Walt Whitman Archive, will also serve as a guest instructor at some of the institutes. UNL has agreed to host one institute.

## Curriculum and Work Plan

Each institute would focus on one of the three themes (manuscripts, contextual information, thematic research collections) described below. These themes have emerged in the WWP's introductory workshops and seminars as particularly pressing, and as meriting sustained attention. They are all areas in which the most recent version of the TEI (published in November 2007) provides new or expanded encoding provisions, or in which recent research has offered new insight that is not readily available in existing training events. They are also areas in which the WWP takes a strong interest, and in which the WWP staff are actively engaged in research as the WWP collection moves into P5. We will thus be equipped to teach institutes on these topics not only from a position of expertise, but also from a position of recent, direct practical engagement and decision-making.

We propose to offer each thematic institute three times during the course of three years, for a total of nine events. Each thematic institute will ideally be offered in all three geographic regions, host agreements permitting.

## Manuscripts

The encoding of manuscript materials is a frequent topic for questions in introductory TEI workshops. Participants frequently bring manuscript materials to practice on, and the number of scholarly projects (large and small) focusing on manuscript collections is growing steadily as these materials are digitized. The manuscript-focused institutes would address issues and challenges surrounding manuscript encoding including:

- How, and at what level of specificity, to represent textual positioning, directionality, and orientation on the page.
- How to represent variant letterforms and their significance
- How to approach regularization (of spelling, punctuation, usage) in the different disciplinary contexts where manuscript encoding is practiced
- How to represent the physical details of the text, both in the transcription itself and also in a manuscript description, using the TEI's new <msDesc>

## Contextual Information

The encoding of contextual information includes information about people and places mentioned in the text, and interpretive information or commentary supplied by the scholar. This is another frequent topic for questions in our introductory workshops, and also a growing area of interest in the TEI community following the publication of the new version of the TEI Guidelines, which includes extensive provision for representing these kinds of data.<sup>3</sup> The issues to be addressed in these institutes include:

- How much information to capture and how to make adequate use of controlled vocabularies to constrain terms
- The logistics of creating and using contextual information, including questions of where the data should actually be stored
- Opportunities for use of (or linkage to) existing authority records

## Thematic Research Collections and Scholarly Editions

The development of large thematic research collections and scholarly editions, in which distinctive issues of scale, data architecture, and work management arise. Many of the participants in our introductory events are contemplating the development of a substantial collection of texts rather than focusing on single texts, and this emphasis is reflected in the realities of funding as well; large collections offer greater impact for the reader and the opportunity for greater intertextual connection. These three institutes will address the following challenges:

- The development of consistent, adequate metadata (including the use of METS)

---

<sup>3</sup> See in particular Chapter 13 of the TEI Guidelines, "Names, Dates, People, Places" (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>).

- The identification of textual features that will provide the greatest benefit to the reader and the most functional basis for interface development
- Schema development and maintenance
- Development and maintenance of consistent documentation
- Opportunities for collaboration with institutional repositories

The curriculum for all of the institutes will be a combination of presentations, discussion, and hands-on practice. In our introductory workshops and seminars, there is typically an even balance between presentation and hands-on work, because there is considerable information to be imparted and participants need an opportunity to experiment on their own. For these proposed institutes, group discussion will play a greater role. This is partly because the participants can learn a great deal from other participants' projects, but also because this discussion constitutes, in a sense, a "hands-on" process which models the real decision-making and project-building work that takes place in digital humanities project development. During these discussions, participants will be able to compare strategies, consider various approaches to the same problem, and gain further familiarity not only with the mechanics of text encoding but also with its intellectual processes. Through examination of case studies, the group will review and discuss the variety of encoding methods being adopted by existing projects (for instance, varying approaches to regularization or to the use of METS), and consider their merits, disadvantages, and overall suitability for participants' own projects. The instructors will present examples and methods, and guide the discussion so that each participant's project receives attention from the group. During the hands-on portions of the institute, participants will have the opportunity both to work on developing encoding rationales based on what they have learned, and also on putting those ideas into practice by working on sample texts from their own projects. A detailed sample curriculum is included in the appendix.

The institute events themselves are deliberately comparatively brief. However, the activities of the institutes are not limited to the physical meetings, but extend beyond these events to offer participants the opportunity to deepen their understanding of what they have initially learned, and solidify it through extended practical implementation. In addition to the costs of travel and teaching, this proposal also funds an important set of followup activities to support participants' ongoing work and to provide consultation both with WWP staff and among the participants themselves. We will offer a combination of structured and unstructured activities both before and after each event:

- In the month preceding each institute, we will ask participants to send us a project description and a set of questions or problems they seek to address at the institute. The WWP staff will review each project's materials both to identify common themes and also to prepare any specific materials that may be helpful.
- Also in preparation for each institute, we will circulate a set of case study materials and readings to participants, to enable them to do some advance preparation if time permits.

- In the month following each institute, we will expect and strongly encourage all participants to complete a draft encoding rationale and documentation, as well as a sample document, and circulate these to the other participants. The WWP staff will comment on these materials and the other participants will be encouraged to comment as well; a listserv will be provided for these and any other discussions. Completion of this work will be optional and without deadline, but we expect that participants will find it useful enough to be worth attempting. In most cases this work will be of direct value to the project itself, and may represent revision or completion of documentation already begun.
- The WWP staff will be available for consultation on the discussion list and more privately via email, to provide advice and technical support for projects during the three years of the grant.
- The curricular materials and any completed samples and documentation will be made publicly available at the WWP web site, so that people who were unable to attend one of the institutes can still benefit from them.

## **Selection of Participants**

A call for participation will be issued at the start off this grant, and will be posted both to the major digital humanities lists (TEI-L, Humanist, CenterNet) and to the hosting institutions for circulation. Applicants will be asked for a brief description of their project (one they are currently engaged in or one they plan to start during the institute) and of their expertise level in text encoding. Applicants will be selected as follows:

- Preference will be given to those whose projects have the highest relevance to the institute topic (and hence will be most fruitful inclusions from the viewpoint of group discussion and comparison). If space permits, however, participants with more marginally relevant projects will be admitted.
- Preference will be given to applicants who have prior familiarity with text encoding. If space permits, novices will also be admitted, with the understanding that they may need to do some advance reading and preparation.
- Preference will be given to collaborative groups (e.g. between faculty and graduate students, or faculty and digital practitioners), so as to enable them to attend the institute together.
- Among applicants with equal weights on the points above, selection will be on a first-come, first-served basis.

## **Impact and Evaluation**

We anticipate several important results from this program.

- It will provide an opportunity for scholars and small projects to get focused, advanced consultation on specific areas of text encoding with strong impact on the scholarly quality of their digital work.
- By treating text encoding not as a purely technical skill but as a problem of strategic design, these institutes will help to raise the level of discourse and the scholarly awareness of text encoding within digital humanities scholarship.

- The visibility of the curricular materials, and of participants' samples and documentation via the WWP web site, will also make the results of this program accessible to a wider audience beyond those able to participate directly.

As described above, a number of followup activities will provide opportunities for participants to continue working on the projects they begin at the institute and receive ongoing guidance and advice from both the instructors and the other participants:

- A listserv will be maintained by the WWP for discussion and questions (for all participants).
- The institute pages at the WWP web site will provide a space dedicated to posting samples and linking to documentation, in addition to the curricular materials developed for these institutes.
- Participants will be strongly encouraged to post sample encoded texts and completed documentation to share with other participants (and will be reminded of the community value of such examples).

We plan several methods of evaluation for these institutes, both during the term of the grant and at its conclusion:

- Participants will be asked to complete an online survey immediately following each event. Experience with this approach in our current grant suggests that response rates are sometimes low despite enthusiasm from participants, so we may ask participants to complete the survey during the last lunch break.
- We will conduct an additional survey at the end of the entire series, to get feedback on the usefulness of the followup consultation and discussion.
- The final report for this institute series will present the projects that participated, and discuss the lessons learned and the impact the discussion and consultation had on significant points in the project design. It will also discuss the value and drawbacks of this approach to advanced TEI instruction.

## **Staff, Faculty and Consultants**

**Julia Flanders** is the Director of the WWP and will be the principal instructor for the institutes. She will lead the design of the curriculum and the creation of supporting materials, and participate significantly in the followup discussion and consultation. She has a PhD in English Literature and has served in positions of leadership in the TEI and digital humanities communities for nearly ten years, and has also taught successful TEI workshops and seminars for ten years. In addition to extensive expertise in TEI encoding, she also has deep knowledge of digital project design and management, documentation strategies, and fundraising.

**Syd Bauman** is the WWP's Senior Programmer/Analyst and will be the second main instructor for the institutes. He will participate in the design of the curriculum and the creation of supporting materials, and will contribute significantly to the followup discussion and consultation, particularly on technical topics such as schema design and automated processing techniques. He has a BA in Political Science and served as North American Editor of the TEI Guidelines from 2001-2007. He has extensive expertise in TEI



schema design and customization as well as TEI encoding, and has been teaching TEI workshops and seminars since 2004.

The **WWP Project Manager** (search is currently under way) is a position requiring extensive familiarity with TEI encoding and with issues of digital project design and management. The WWP's activities during the coming year will emphasize the development of structured contextual information (using the TEI's personography and toponymy encoding), and the Project Manager will have significant expertise on this topic by the time the institutes are being taught. The Project Manager will contribute to the followup consultation and discussion and add breadth of perspective to the information and advice available to participants.

**Dorothy Carr Porter** is the Program Coordinator for the Collaboratory on Research in Computing for the Humanities at the University of Kentucky. She has an MA in Medieval Studies and an MLS, and significant expertise in manuscript encoding, having served on the TEI Council and the TEI's manuscript description workgroup, as well as on the executive board of *Digital Medievalist*. She is also the chair of the Committee on Electronic Resources at the Medieval Academy of America. She will serve as a guest instructor at the institutes whose emphasis is on encoding manuscripts, and as local organizer for the institute being held at the University of Kentucky.

**Brett Barney** is Research Assistant Professor at the Center for Digital Research in the Humanities at the University of Nebraska and Senior Associate Editor for the *Walt Whitman Archive*. He has a PhD in English and extensive experience with TEI encoding, particularly with respect to manuscripts, metadata, and thematic research collections. He has designed encoding for various electronic editing projects, including *The Journals of the Lewis and Clark Expedition Online* and the *Willa Cather Archive*.

A third guest instructor will be selected if the institute is funded, with expertise in metadata and the management of contextual information.

## **Budget Notes**

The budget for this proposal covers salary and benefits for the two WWP staff who will be serving as instructors for the institutes, plus a small fraction of time for the WWP project manager who will contribute to the consultation and discussion in the followups to the institutes. We include funding for one guest instructor per institute, with an honorarium of \$1500 covering the three days of the institute plus two days' preparation. Travel funding is provided for the three instructors. Funding is also provided to cover travel funding for participants: up to 10 participants per event, at approximately \$500 per participant. \$2000 per event is also provided to cover the basic hosting costs of room and equipment rental, and also to provide lunch and coffee breaks for participants.

## ***Lunch break***

### ***Session 3: Document analysis***

Participants will complete a detailed document analysis of their sample text(s), following guidelines provided by the instructors. The document analysis will help participants to identify the significant textual features, and also the significant functional desiderata, that will motivate the design of their project's TEI encoding approach.

### ***Session 4: Hands-on***

During this first hands-on session, participants will begin work on a sample text, with the goal of creating an initial light encoding of a short representative section. The instructors will work one-on-one with the participants to answer questions and suggest encoding approaches.

## **Day 2**

### ***Session 5: TEI Customization***

Introduction to TEI customization using the TEI's web-based schema customization interface (Roma). Participants will learn both the motivations for TEI customization (which is essential for any advanced work with the TEI, since it permits the selection of elements and the constraint of attribute values) and the essentials of creating a TEI customization.

### ***Session 6: Hands-on***

During this second hands-on session, participants will create an initial customized TEI schema, responding to the needs articulated in their document analysis, and will revise and extend the encoding of their sample document based on the schema's constraints and provisions. The instructors will work one-on-one with the participants as needed to assist with the customization process.

## ***Lunch break***

### ***Session 7: Discussion***

This discussion session will focus on the problems and themes that have emerged from the hands-on encoding thus far. The instructors will structure the discussion around the most common challenges and suggest encoding strategies for the group to consider. The aim of the discussion will be both to help participants arrive at possible encoding solutions, and also to raise participants' awareness of the wide variety of encoding approaches, and the conditions under which each one might be appropriate.

### ***Session 8: Encoding strategy development***

During this hands-on session, participants will be given an opportunity to begin outlining a document describing the encoding strategy for their project, giving details of specific encoding problems and approaches. This document will serve as the framework for more extensive project documentation to be completed after the institute. Participants may also continue working on refining their schema customizations and the encoding of their sample document, based on

insights deriving from the previous discussion. Instructors will work one-on-one with participants as needed to provide guidance and assistance.

## **Day 3**

### ***Session 9: Case studies***

This session will focus on the examination of sample texts and documentation from a group of relevant projects. Participants will be encouraged to review these materials in advance, and in the discussion the group will consider the applicability of specific strategies to the participants' own projects.

### ***Session 10: Final presentations***

In the final three sessions, each participant (or project group) will have the opportunity to give a short presentation (about 15 minutes) on their project, with emphasis on the encoding strategy and customization work they have done during the institute, and on how these respond to specific challenges and functional desiderata. Each presentation will be followed by questions, critique, and discussion by the group, led by the instructors. In all there will be approximately 6 hours of time allocated to these presentations, or about half an hour per person or project group.

### ***Lunch break***

### ***Session 11: Final presentations***

As above.

### ***Session 12: Final presentations***

As above.