

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.³ 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)

³ 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>).

- 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.

Institute Curriculum

Advanced Topics in TEI Encoding

Day 0 (optional)

This optional first day would be provided if a significant number of the participants had no familiarity with the TEI, or if a one-day introductory TEI session proved useful for the local organizers.

Session A: Fundamentals of XML

Coverage of essential concepts of XML markup languages, including elements, attributes, schemas, well-formedness, and validity.

Session B: Fundamentals of TEI

Introduction to the most commonly-used TEI elements, using extensive examples

Session C: Hands-on

The first afternoon session will be spent in hands-on practice so that participants can get a basic grounding in the essentials of TEI markup

Session D: More advanced encoding concepts

The final session on the first day will be spent quickly introducing concepts of linking, regularization, parallel structures, and metadata.

Day 1

Session 1: Introductions

Initial presentation of projects: participants give brief descriptions of their project materials and goals

Session 2: Topical fundamentals

Presentation and discussion on topical encoding fundamentals (manuscripts, contextual materials, thematic research collections), introducing the essential TEI features that are germane to the institute's specific encoding domain. The instructors will present encoding concepts and examples for the group to consider, and the discussion will focus on identifying features that are likely to be most useful to the participants, and features which may require further thought.

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.