White Paper

NEH HD-248437

TEI and Humanities Pedagogy: Building TAPAS Classroom

Project Director: Julia Flanders

Northeastern University

April 23, 2018

Introduction

This white paper reports on work completed under an NEH Digital Humanities Startup grant that ran from May 2016 through October 2017.

TAPAS (the TEI Archiving, Publishing, and Access Service, http://www.tapasproject.org) offers a repository-based framework for scholars to publish and archive their TEI-encoded materials, including scholarly editions, text anthologies, digitized archival materials, historical papers, manuscripts, and a wide range of other TEI data. TAPAS was founded to address the challenges faced by producers of TEI data who, having produced richly marked texts representing intensive digital scholarship, need a way to disseminate that work and make it visible. TAPAS seeks to make it possible for scholarly practitioners of all kinds—even those without access to technical resources—to be creators and curators of high-quality data, by providing a powerful framework for the preservation and publication of TEI projects. TAPAS also provides the technical expertise and long-term stable base of curatorial support needed to maintain meaningful access to these projects, even after their individual scholarly creators can no longer do so. TAPAS was founded by a multi-institutional collaboration led by Wheaton College and Brown University, and is currently hosted at the Northeastern University Library's Digital Scholarship Group.

In recent years the number of resources available for scholars to learn and adopt the TEI has grown: workshops, self-teaching guides, publication tools, and endorsement by professional bodies and funding agencies. As a result we now see a strong interest among humanities scholars in using these tools in their own teaching at both the graduate and undergraduate levels. The Women Writers Project at Northeastern University recently completed a well-attended series of advanced TEI workshops which included "Teaching with TEI." As those workshops demonstrated, the real challenge for teachers is not in explaining TEI, but in providing the concrete environment for a TEI classroom: one in which text markup can be studied from different perspectives, compared, annotated, and viewed in useful ways on the web. The logistics of classroom encoding assignments can prove challenging: tools for displaying and publishing XML-encoded data typically require greater technical expertise than humanities faculty possess, and these tools are not designed with classroom needs in mind. The limitations that many instructors face in both time and technological resources can be a barrier to these assignments, even in cases where there is strong interest from both students and instructors.

The TAPAS Classroom initiative funded by this grant builds on the TAPAS project and seeks to address these limitations by creating an extension of the TAPAS framework aimed specifically at supporting the use of TEI in the humanities classroom, including traditional humanities and digital humanities courses, workshops, and pedagogical involvement of students in digital humanities projects. Working with a small group of active TEI instructors, the TAPAS team developed a specification and road map for TAPAS's planning for pedagogical support, developed visualization tools aimed at and inspired by classroom use, developed tools for user-friendly validation and consistency checking, created a set of documented templates and sample files to help instructors scaffold TEI classroom activities, and built a supporting framework for repeated events such as workshops. The development work was completed and operational on the TAPAS development server as of the time of writing, but not yet visible to the wider public because of the delayed completion of infrastructural work on the underlying

TAPAS systems, resulting from the departure of the Digital Scholarship Group's web applications programmer. The anticipated date of full release is late summer or early fall 2018.

Project Activities

Pedagogical Spaces in TAPAS

We began the grant period with intensive discussions (both within the TAPAS team, and with our project advisors) of the ways in which we anticipated TAPAS being used in pedagogical contexts. As part of these discussions we considered the kinds of additional documents teachers and students would need to upload and manage within TAPAS, the kinds of communications between teachers and students that the system would ideally need to support, and the kinds of viewing options for TEI data that would be most fertile for teaching TEI.

From these discussions, we developed a detailed specification for the Classroom component of TAPAS. This specification also includes a road map that describes features that are outside the scope of this grant award, but which we hope to come back to in a future development phase of the project.

The central features we decided to focus on in the Classroom expansion of TAPAS are:

- Expanding the range of stylesheets and viewing options as a way of helping students explore the effects of markup on textual presentation, recognizing that these viewing options help reinforce students' understanding of the relationship between markup and the many different kinds of output it can generate.
- Providing for user-friendly validation of TEI files, to support classroom discussion of validity and the relation between the TEI data and the schema(s) within which it is in dialogue
- Providing a set of sample documents and templates to serve as starting points for encoding and as examples of template-building that faculty could adapt to their own classrooms
- Providing a space where instructors can share pedagogical materials such as sample files, templates, and schemas

The features which remain on the road map for the future include:

- Supporting the upload of other types of supporting materials including syllabi and assignments
- Supporting validation against user-provided schemas
- Supporting user-provided CSS (this is already supported to some extent, in that in one of the view packages the text display reflects the CSS and other renditional code embedded in the markup)

View Package Development

One key goal of the project was to create new document views within TAPAS that are specially aimed at pedagogical use, both to take advantage of the semantic information captured in TEI

markup, and also to facilitate study of the markup itself. To support this goal, we developed several new viewing options within TAPAS, taking advantage of the underlying architecture in TAPAS by which TEI data is made visible. To display TEI data on the web, the TEI code is typically transformed into HTML using stylesheets written in the Extensible Markup Language for Transformations (XSLT), usually accompanied by Cascading Stylesheet (CSS) code that specifies the formatting. TAPAS's display of TEI data embeds XSLT and CSS stylesheets in a larger "view package" that also contains JavaScript to support user interaction, a manifest that lists and documents the components of the view package and their functions (to enable them to be easily modified and reused), and an optional schema that can be used to test the target TEI data to determine whether it is appropriate for use with the view package. TAPAS view packages take a standard form (described by the manifest) and are maintained in GitHub; the TAPAS architecture includes a regular mechanism by which the system checks the current inventory of view packages, detects new and updated packages, and updates the TAPAS interface accordingly. Adding new view packages is thus comparatively straightforward and does not require knowledge of TAPAS's internals, making it possible for TAPAS to work collaboratively with the TAPAS user community to develop new view packages for specific purposes. (This architectural work was performed under an NEH Research and Development grant that was recently concluded.)

For TAPAS Classroom, we developed four view packages with specific relevance for pedagogy. In the first of these, we completely overhauled the view package called "TAPAS Generic", which is the viewing option that is designed to display the broadest range of TEI data, and we dramatically expanded its functionality in ways that responded to requests from TEI instructors and faculty. In particular, we focused on enabling users to control the display of the most commonly-used phrase-level features in TEI so that students can see the results of their encoding work through the specific editorial choices expressed in the display. We added the ability to toggle between diplomatic and normalized views of the text (taking advantage of TEI markup that captures abbreviations, old spellings, and typographical peculiarities of the text), enabling students to create documentary editions of archival source documents and see the impact of their markup upon the text. We also added support for contextual information (biographic, geographic, and bibliographic data about named entities in the main TEI texts). In documents whose encoding includes explicit identification of named entities (e.g. using <persName>, <placeName>, and so forth) and linking to a managed list of those entities (using <listPerson>, <listPlace>, <listBibl>, and so forth), the view package will display those entities with a mouseover showing the individual entry with biographical or geographical or bibliographic information. The package also displays the entire authority list as a document that can be browsed in its own right. There were several challenges in presenting data of this kind. First, it is potentially quite open-ended, since different TEI projects record differing levels and types of descriptive detail, so a viewing option needs to gracefully scale both up and down to very simple records and very complex ones. And second, the formatting of the authority list needs to support easy skimming and navigation from record to record (potentially over a very large number of records), providing clear visual cues to the specific types of information provided, as well as good navigational support. This new view provides clear formatting for all potential types of data in these contextual files and intelligent fall-back options to handle the variations within the data.

The second view package for TAPAS classroom is titled "Hieractivity", and the name of this view package conveys its two key display concepts: the hierarchy represented by the TEI document's encoding, and the forms of interactivity this enables for the user. This view package was developed to give students a way of seeing the overall encoding structure of a document, in manner similar to the spirit of "distant reading", and it uses color coding to make the distinct layers of the document visible and intelligible. It provides the ability to toggle certain aspects of the highlighting on or off (so that the reader can focus on specific elements of interest), and it also provides information about the encoding (including the raw counts of each element). As a result, it offers interesting teaching possibilities: for instance, the ability to compare the encoding of different documents, or to get a high-level view of the overall prevalence or distribution of a specific element. If multiple students have encoded the same document, this view can prompt a fruitful discussion of encoding decisions and where practices diverged and converged. This view also gives students a different view of the encoding in which the structure of the XML is more intuitively evident (which can be helpful for those who are still uncomfortable with XML notation).

The third view package is aimed at showing students an "under-the-hood" view of the TEI data. This view package displays the raw XML, in a form that allows elements to be folded and expanded for convenient navigation. The presence of the XML view enables students to move between the formatted views and the underlying markup, to understand more clearly how the markup is influencing the various forms of formatting and to support trouble-shooting.

The final view package focuses on XML validation, which is the process of testing a specific XML file against the schema that defines the specific markup language it uses (in this case, TEI). Validation plays a crucial and complex role in TEI pedagogy and in TAPAS. For the TAPAS Classroom project, we wanted to make validation a less mystified and more helpful process for students, getting past a "parser is mad at me!" level of perception. We are now testing a pedagogical validation display that operates as a view package, displaying a report of user-friendly validation errors and advice on how to address them. Treating validation as a view package is an interesting variation on the concept of the view package: it shows that the display of TEI/XML data need not be "documentary" in the sense of reproducing the original content and order of the TEI file; it can represent instead an analysis, an alternate view, a report, a set of derived information. This approach works for validation but it could also be extended for other kinds of report generation (for instance to show text statistics, or to graph the connections between texts). This view package displays a validation report on the text, with links from the error messages to the text itself. The view package currently validates the text against the TAPAS schema, which is close to the TEI's most inclusive schema. Future versions of this view package will allow the user to choose from a menu of different schemas (eventually including user-contributed and project-specific schemas) and also from a variety of other forms of consistency checking using tools such as Schematron. The view package takes the output of a validation process (run within the TAPAS XML database) and compiles it together with the TEI-encoded text, and then converts the whole into HTML for display. It also maps the original error messages from the validation output (which can be somewhat technical) onto user-friendly versions written by TAPAS, for greater comprehension by users. The goal is to provide a report on the text that is framed not as a forbidding set of "errors" but rather as helpful information about how the encoding matches up with different available schemas; TAPAS does not want to position its schema(s) as *authoritative* but rather as *informational* with respect to

specific contexts of usage. A version of this view package will be ready for release by the time the new TAPAS site is released, hopefully in summer 2018.

Pedagogical Scaffolding

Another important goal of TAPAS Classroom was to offer instructors a centralized and user-friendly platform for organizing and sharing course materials, with features to support group analysis, display, and commenting on TEI assignments. We developed several support structures for instructors, both to facilitate the use of TAPAS in workshops and repeated courses and also to scaffold encoding assignments. First, we developed a set of commented templates and sample documents in various common genres (such as letters, poems, prose documents) that illustrate the basic markup involved and provide a simple starting point for classroom encoding activities. These templates are useful to instructors who lack the time or the expertise to develop templates of their own, and they can be revised and adapted to specific classes or assignments. Templates provide scaffolding for encoding exercises by ensuring that students start with a valid or close-to-valid TEI file, which they can then modify as they transcribe content into the template's skeleton structure. Sample files provide the opportunity to learn by example, so that students can see markup in use.

We also created a shared TAPAS Classroom space (the Community Workshops project) in which templates and sample files are organized, and within which instructors can create workshop-specific collections. The goal of the Community Workshops space is to reduce the administrative overhead for workshop instructors and teachers who want to make brief use of TAPAS in a pedagogical context. We considered building this feature set into the individual project space and decided against it; after further discussion within the project we realized it was important to distinguish between making TAPAS a pedagogically-attuned TEI publication/archiving space, and making TAPAS into a course management system like Blackboard, which lay outside the scope of our effort and intentions. Instead, we wanted to provide ways for instructors to share materials that were effective, and make it convenient to set up and use TAPAS projects in pedagogical contexts (rather than treating TAPAS itself as a general-purpose pedagogical space).

Audiences and Impact

TAPAS Classroom was designed with several main audiences in mind. The most important of these is humanities faculty who are using TEI in their courses, and the students in their classes. This is a growing area of need, both because faculty who use TEI in their research are increasingly finding ways to make connections between that work and their teaching, and because faculty are being encouraged by their institutions to explore digital humanities pedagogy. Overall, there is an increasing demand for courses that build bridges between humanities and more quantitative or formal literacies. Having a framework like TAPAS enables faculty and students in these classes to focus on the TEI encoding enterprise without having to carry the extra burden of maintaining a TEI publishing environment. A closely related audience is teachers of TEI workshops—shorter events focused solely on TEI, often repeated in the same format. Here, the convenience of using the TAPAS framework is important, but equally so is the opportunity to introduce workshop participants to TAPAS as a possible venue for their nascent

TEI projects. The question of how to publish TEI data naturally arises at the end of an introductory TEI workshop, so TAPAS is a good way for instructors to address it without having to teach additional skills such as XSLT.

A third important audience is IT and library staff who provide support to humanities faculty, or work in collaboration with humanities faculty on courses that use TEI. In our small sample of courses that use TAPAS Classroom, we had two participants in this role, a significant proportion. Staff in these roles are an important gateway to TEI expertise and supporting tools, and the availability of a service like TAPAS can help them make the case for using TEI, by reducing the barriers to entry and simplifying the process to a manageable level.

A final audience of a different kind is scholars interested in DH pedagogy: in the tools and platforms that support the teaching of digital methods, and in their impact on the teaching and learning process. TAPAS Classroom is an important intervention in this domain, not necessarily because it provides an optimal approach, but because it addresses a set of common challenges and proposes a particular kind of solution which others can learn from and build on. As the instructor of one of our test courses observed in an article about the course, for liberal-arts colleges where digital scholarship has its greatest traction in pedagogy, one key challenge is how to provide effective technical infrastructure for courses taught by humanities faculty, without requiring them to become experts on that infrastructure. The approach TAPAS Classroom proposes is to centralize that infrastructure and take advantage of its alignment with research uses of digital tools, rather than trying to support pedagogy as a completely separate activity. As the TAPAS Classroom initiative demonstrated, in some ways this alignment is a definite strength and may provide an added dimension to courses, but it also complicates the design and development process significantly. The question of whether DH pedagogy can naturalize within tools of this kind (and we might here also consider similar tools like Omeka, Mukurtu, TextGrid, CWRC, and other scholarly digital content management and publishing systems) or whether it requires separate, pedagogically-specific spaces, is an ongoing research question for which TAPAS is a relevant test case.

Because of the short duration of the grant period and some staffing complications we experienced, this project was focused on development and preliminary testing; the results have not yet been publicly rolled out. As a result, the impact on these intended audiences has not yet fully unfolded. Our preliminary testing was focused on a set of case studies representing a range of different scenarios, mapping onto the audiences described above. The feedback we received from those test cases was valuable both in confirming that TAPAS Classroom (even in its still-buggy and preliminary form) met the needs of those scenarios, and also in providing guidance on how to shape the service even more fully in the directions they needed. This process also revealed an unforeseen audience element: we had not anticipated that undergraduate students involved with faculty-led TAPAS projects might become research partners and project leaders, but we learned of at least one case where this happened. This both confirms our sense that pedagogy and research uses of TAPAS are very tightly connected, and also suggests that there may be further functional improvements to better support this kind of role transition (for instance, in the ways user roles are defined).

One impact we are hoping for from this initiative, but have not yet been able to measure, is increased TEI membership from TAPAS Classroom users. In particular, we anticipate that faculty may request that their institution join the TEI Consortium as a way of getting access to

the TAPAS platform to support their teaching, and we also anticipate that TEI workshop participants (who receive TEI membership free for one year as a benefit) may extend their memberships after the initial free year if they find TAPAS useful after being introduced to it in the workshop. We are working with the TEI Consortium on ways to track this impact.

Evaluation of Outcomes

The strengths of the project clearly lie in the TEI and XML expertise of its development team. Excellent work was done in designing view packages that work with very complex and unpredictable TEI data, and that also do revealing, interesting, and pedagogically useful things with that data. Part of this design process also involved very thoughtful and intensive work on the deeper architectural aspects of these packages. This project required us to think very carefully about details like how the components of the view packages would be inventoried and described, how the view package system could be documented to prepare in advance for community-contributed stylesheets. Most of this work is not directly visible in the user interface but has a longer-term impact on the maintainability of the view packages (as TEI data changes) and also on our ability to involve TEI community members in creating new viewing options. Our attention to concerns of longevity and sustainability is consistent with TAPAS's overall aims of publishing and preserving TEI data for the very long term. The web development aspect of this project was a bigger challenge and also revealed a vulnerability in the DSG team (and hence the TAPAS team). As noted above, web development skills are challenging to staff in any academic environment because of competition from industry, but in addition, this project required more specialized skills (including familiarity with digital repository frameworks) which made it much more difficult to replace an outgoing staff member.

At this stage, having completed but not fully released the TAPAS Classroom features of TAPAS, we see the project as successful at addressing the challenges of TEI+pedagogy from the perspective of how TAPAS supports TEI usage in the classroom, but we see unfinished lines of inquiry and more work to do on the pedagogical side: we expect to learn a great deal more once a wider group of teachers have put TAPAS into use in their classrooms. However, the case studies with our initial group of faculty showed that the service does meet an important need. Their enthusiasm for the tool and their sense of its further potential are very encouraging. As noted earlier in this report, there is a scoping challenge here, in that we are not seeking to make TAPAS into a learning management system or a course management system, although we can anticipate that some of the needs and requests we receive from faculty users may lead in that direction. The design challenges of skirting that border—providing strong support for classroom use without making TAPAS a self-sufficient online classroom space—are significant but one important outcome of this project was that the border is now much clearer to us.

Continuation of the Project

TAPAS is intended as a very long-term project and has the support of the Northeastern University Library and the Digital Scholarship Group as well as the TEI Consortium. We plan both general continuation of the project as a whole, and also specific attention to the goals of TAPAS Classroom as initiated under this grant. Specific goals for the near term include:

- Complete the validation view package, which is very close to being completed
- Finish work on the overhaul of TAPAS as a whole so that we can roll out the TAPAS Classroom features
- Complete that roll-out and publicize new features; recruit additional faculty and workshop instructors
- Work with the coordinators of the TEI workshops for the Digital Humanities Summer Institute (University of Victoria) on integrating TAPAS Classroom into DHSI TEI workshops; we hope that this will be the start of a long-term relationship with DHSI that can model the value of TAPAS for other workshop and institute programs

In the longer term, there are further goals arising from TAPAS's ongoing work:

- With the hire of new web applications developer at DSG, we will address several crucial features including establishing a public API to TAPAS data, and providing support for embedding a TAPAS frame in another publishing framework such as a WordPress or Omeka site.
- We are planning to seek funding to formalize and publicize mechanisms for community members to add new view packages, and to work with specific segments of the TEI community to design high-value view packages for specialized types of data such as scholarly editions.
- As the Digital Scholarship Group's infrastructure for digital scholarly publishing matures, we will explore ways of bringing TAPAS into that framework, which would enable us to support additional features such as web-based XML editing.

Long Term Impact

We anticipate several forms of long-term impact from the TAPAS Classroom initiative. First, we expect to see increased usage of TAPAS in courses and workshops, with more frequent integration and more effective integration of undergraduate and graduate students into TEI projects. Because TAPAS Classroom offers enhanced ability to teach TEI with better scaffolding (e.g. user-friendly validation, encoding samples and templates, and the ability to see immediate published results), we also expect faculty to have a more positive experience integrating TEI into their teaching, with less start-up effort. As a derived outcome, we hope to see the emergence of a community of TEI-engaged humanities instructors who use TAPAS regularly and can provide mutual support and exchange of ideas.

At a broader level, any increased integration of TEI and concepts of text markup into humanities teaching also tends to reinforce the value of data modeling as part of humanities and digital humanities education; in a world where so much of our social, political, and cultural interaction is digitally mediated, the shaping of data ("content") has a profound effect on how information circulates and as such is a core aspect of literacy. TAPAS Classroom immerses students in the data modeling process and its outcomes, giving them power over how data is created and insight into the interplay of semantics, presentation, and usage, as well as practical experience with digital publishing workflows. It also offers support for a version of digital humanities that is not solely preoccupied with large-scale data and programming, but rather with modeling as a tool for thinking analytically, critically, and interpretively: reinforcing rather

than de-emphasizing these specifically humanist competencies. Future features of TAPAS Classroom—such as community-contributed view packages—will further explore how seeing texts differently can reinforce textual engagement. Overall, TAPAS Classroom helps to position the work we ask students to do with primary sources not just as a kind of facsimile creation, but also as a way of creating dynamic and responsive versions of the text in which they are creators as well as consumers.

Grant Products

Several grant products were produced during the grant period and several others will be finalized and disseminated in the coming year:

- A plenary presentation was given at the TEI 2017 conference. Slides and notes are included in the appendix and the abstract is visible at the conference site: http://hcmc.uvic.ca/tei2017/abstracts/t_137_flanders_tapasclassroom.html.
- An article based on this presentation is being submitted to the Journal of the TEI Consortium (https://jtei.revues.org) and if accepted will be published in issue 11 with other selected papers from the TEI 2017 conference.
- Enhancements to the TAPAS site will become visible during summer 2018 at http://tapasproject.org

Materials from several courses in which TAPAS was used and tested are available online:

- Professor Karen Bourrier's "Digitizing Victorian Women Writers" course web site (https://digitizingwomenwriters.wordpress.com) and project code book (https://docs.google.com/document/d/19SQ84BYuegc7KB-LEe1hQL4WinFlzH4ljq4Zsa2wwHA/edit)
- Professor Kristen Bennet and Scott Hamlin's "A Rogue's Progress: Mapping Kit Marlowe's Social Networks" publicly visible course materials in TAPAS (http://tapasproject.org/kit-marlowe/marlowe-related-texts) and supporting materials used for the project (https://kitmarlowe.org/mini-archive/schemas-and-odd-files/)

The source code for TAPAS view packages and other components developed or enhanced under this grant is available via GitHub:

- Code for the TAPAS Generic view package: https://github.com/NEU-DSG/tapas-generic
- Code for TAPAS view package architecture: https://github.com/NEU-DSG/tapas-view-packages
- Repository of TEI samples and templates, plus the code that generates sample and template files from a commented source: https://github.com/NEU-DSG/tapas-TEI-files