

Data Management Plan

As part of the TAPAS project, TAPAS Classroom's data management planning is framed within the expectations of TAPAS itself. TAPAS is a long-term, repository-based publication framework for TEI data, and in addition to serving the data curation needs of its member projects, it also serves as a repository of TEI data for the TEI community as a whole. Because the TEI itself is designed with the goal of supporting long-term data curation (and has already been in wide use for over 20 years), TAPAS project staff is deeply committed to supporting that goal by managing and curating community created TEI and data.

All data submitted to TAPAS for archiving is stored in a Fedora/Hydra repository hosted by Northeastern University Library. This hosting relationship is governed by a memorandum of understanding which stipulates that the Northeastern University Library will provide long-term storage of and access to unlimited quantities of TEI-encoded data for 20 years at no cost, with the option to extend this period by mutual agreement; Northeastern also provides storage of non-TEI data (including image files) on a cost-recovery basis. The TAPAS repository is maintained by Northeastern University's Library Technology Services staff and Information Technology Services staff on the same basis as Northeastern University Library's own Digital Repository Service (DRS), with respect to backups, data integrity checks, and other storage and preservation practices.

TAPAS Classroom data falls into three categories. The first category is TEI files uploaded to the TAPAS "sandbox" by users with free accounts (including students), which will generally be simply test data: for instance, TEI files uploaded to see how they look on the web, or as part of a short-term assignment. To avoid encumbering the TAPAS repository with ephemera, we will provide an option for users to identify such files as test data, in which case such files will be deleted after a short period (to be determined in concert with our test community), and the standard "test" workflow will alert users to the short-term retention. Users will also receive an email alert prior to the deletion, with the option to move the data into the TAPAS Commons for long-term storage.

Within TAPAS Classroom proper, a second category of data includes course assignments, schemas, documentation, and other supporting files. These will be treated as repository items, and will be stored and preserved according to existing DRS preservation policies. The third category is TEI files uploaded by students within the formal framework of TAPAS Classroom. By default, the work completed by students as assignments will be retained for a period of three years to permit the completion of a degree in which these materials might remain relevant. At the end of three years, individual students may choose to submit their TEI Classroom materials to the TAPAS Commons or to create a TEI membership of their own and migrate their class work into a TAPAS project. They may also download their work and retain it personally.

During the year-end course transition process, instructors may select which parts of the course to carry over to the next course by:

- downloading the entire course unit (including course materials and student content);
- migrating the entire course into a TAPAS project (in which case all student materials would be treated as repository items); this option would be appropriate if the class collaborated on a project such as a digital edition; or
- retaining the course framework for reuse while clearing out the student assignments.

Additional options for managing TAPAS Classroom data may be added during development of the system specification.

All TAPAS program code, including any future code developed for TAPAS Classroom, is maintained in a GitHub repository under an open-source license.