

7. DATA MANAGEMENT PLAN

The Tesserae Project's products fall into the following categories.

- **Website.** The publicly-accessible front-end to the tool, including its interface and documentation.
- **Code.** The scripts which parse input texts, create feature sets, perform searches and format results, as well as those which create the web interface.
- **Texts.** The corpus of classical texts which we use and provide via our web interface.
- **Results.** The lists of textual parallels with associated scores which are the output of the search program.
- **Analyses.** The product of philological research performed on specific case-studies using the tool, including aggregated results of multiple searches, statistical studies, and literary interpretation of computational results.

All the data mentioned above will be made available in a timely fashion to the research community and the public, and will be hosted indefinitely by the University at Buffalo (UB). Category-specific details and timelines for availability are outlined below.

Website

The project's website will be hosted indefinitely by UB's College of Arts and Sciences Educational Technology web server.

Code

All the code produced by the project will be owned by the University at Buffalo, which will release it under a free and open-source license. The current stable version of the program will be maintained in a version control server at UB. Information on how to download the code anonymously will be published on our website.

Texts

All the texts used by the online version of the program will be maintained on the UB web server, and publicly accessible via the website. Texts will be stored as TEI P5 conforming XML documents.

As an intermediate stage between text and results, the program will create indexed digests of the texts and store them locally in order to speed up on-line interaction with the tool. These digests will not be shared, since (a) they are large and slow to transfer and (b) they can be recreated using the texts and software we already plan to make available.

Results

The program's output, collated lists of ranked linguistic parallels, are produced on-demand. For web-based users, they are presented as dynamic HTML content, and may be downloaded and saved as an XML document or a list of comma separated values. After the user's session is over, they are not stored permanently on the server.

Analyses

Our own work, on both the development new methods and interpretation of results, will be published initially on the website in the form of white papers. Our goal is to publish our major findings in scholarly journals; to the degree that we are permitted we will retain the material itself or links to online publications elsewhere on the website. We will also provide a bibliography of published work on the website. Data referred to in these studies will be retained on the website. In case future changes to the program cause results to be no longer replicable, we will clearly indicate the version of the program used, both in our reports and in any stored data.