

Data Management Plan for the Start-Up Phase of Annotation Studio

Data to be generated

Type of data	When shared?	Under what conditions?
Open Source computer code associated with tool, interface, and server-side components development.	At conclusion of the start-up project, when initial testing has been completed.	Code will be freely available.
User-generated texts, annotations, and electronic multi-media collections generated during the testing phase.	None of these data will be publically available until the conclusion of the start-up project.	These data will be made available at the discretion of the creator, who will control access via privacy settings. Copyrighted data will be dynamically excluded from collections and documents made available to the public.
Assessment data generated during the testing phase.	Aggregated data will be shared via the white paper and final report to NEH.	No information will be shared that could identify individuals participating in the assessment process.
White paper	After the project has been completed.	The white paper will be freely available to the public via the project website.
A multimedia report posted on the project website	After the project has been completed.	The multimedia report will be freely available on the project website.
A final report to NEH	At the conclusion of the project.	Dissemination of the final report will be the responsibility of NEH.

Period of data retention

Data will be retained for 5 years beyond the completion of the start-up phase of Annotation Studio. Reports and aggregated data will be publicly available within 1 year of project completion, via the project website, copies of the data will be stored long-term in DSpace@MIT. <http://dspace.mit.edu/>

Data formats and dissemination

Computer code will be available as open source in a publicly accessible code repository (GitHub). Reports will be made available in PDF format and disseminated via the project website and through DSpace@MIT. All metadata associated with media documents in shared multimedia collections in Annotation Studio will be freely available on the Annotation Studio website, only

copyright cleared media and text documents will be accessible to the public.

Data management and maintenance

All computer code will be stored in Github, where existing code base for Metamedia (Annotation Studio's precursor) and other HyperStudio projects and tools has been stored. <https://github.com/>.

All other data, including user generated texts and annotations, (exportable as TEI/XML files), assessment data, reports, and publications will be stored in DSpace, MIT's online institutional repository for faculty and researchers. Copyright-cleared media documents will be stored in DSpace along with their respective collection and student interaction information.

DSpace@MIT identifies two levels of digital preservation: bit preservation, and functional preservation. Bit preservation ensures that a file remains exactly the same over time – not a single bit is changed – while the physical media evolve around it. Functional preservation goes further: the file *does* change over time so that the material continues to be immediately usable in the same way it was originally while the digital formats (and the physical media) evolve over time.

DSpace@MIT insures permanent data preservation in a secure and searchable archive, and is managed by MIT Libraries.