

7. Data Management Plan

A. Roles and responsibilities

This data management plan will be implemented and managed by Benjamin Brochstein, with consultation of Chad Shaw. The Rice Digital Scholarship Archive (RDSA, <http://scholarship.rice.edu/>) will have long-term responsibility for the permanent storage needs of all data. All transferred data will be made publicly accessible, any computer code or packages will also be posted on GitHub.

B. Expected data

We are developing a set of parameters and requirements for developing software objects. Therefore, our data is at two levels: the collation of the data into a set of parameters and objects, and the documents from which the data is collated including questionnaires, white board preservation, and workshop notes. The data from preservation of objects will include:

- a comprehensive list of potential uses for DTA
- a comprehensive list of DTA techniques and the advantages and disadvantages of each
- a set of requirements for developing an intuitive user interface to access
- answers to the pre-workshop questionnaire
- photos and/or video of the whiteboard

C. Period of data retention

All relevant data will be deposited in the Rice Digital Scholarship Archive (RDSA, <http://scholarship.rice.edu/>) for long-term storage upon completion of the project study. Once data is transferred to the RDSA, all data will be made publicly available immediately for a period no less than five years. No data will need to be retained for other purposes.

D. Data formats and dissemination

Specifications, requirements, and answers to the questionnaire will be in Word format and posed on the RDSA. The distributable package will be an R package for small corpus DTA tools posted on GitHub.

E. Data storage and preservation of access

All word documents will be deposited in the Rice Digital Scholarship Archive (RDSA, <http://scholarship.rice.edu/>), that has capabilities to manage, archive and share digital content. The RDSA allows the public access to the stored material via persistent URLs, provides tools for long-term data management, and permits permanent storage options. The RDSA has built-in contingencies for disaster recovery including redundancy and recovery plans. The distributable R package will be stored at no cost on the Comprehensive R Archive Network ([CRAN](http://cran.r-project.org/)) that also has built-in contingencies for disaster recovery including redundancy and recovery plans.