

Data Management Plan

Roles and responsibilities

Prof. King-Ip Lin and Prof. Elise King will both oversee the data management plan for the project. Prof. Lin will focus on the software that is developed throughout the project, while Prof. King will focus on the management of the floor plan data, together with the interaction with University of Texas Alexander Architectural Archives for their floor plan data.

Expected data

The following type of data/information will be generated from the project:

1. Source code for BuDAS. A final, working version of the source code of BuDAS will be made available on GitHub at the end of the project. We anticipate BuDAS will be using other public-domain source code. That information, together with the location of the necessary code, will be included in an installation guide that comes with the software on GitHub.
2. Sample data for testing. We will include a set of floor plans as sample data. The data will come in various formats, all of them will be available via GitHub as well as the Texas Data Repository (see section 5):
 1. Original floor plan images (PDFs)
 2. Data that is transformed by BuDAS into the database (stored as a text file containing SQL statements to construct the database)
 3. Ground truth: all the information that is supposed to be captured (e.g. room relationships for each house) will be stored in a text file
3. White papers, reports, and publications will be made available through a web page dedicated to the project. The page will be managed by the Project Directors as well as the system support staff at the Department of Computer Science at Baylor University. We will also make these items available via the Texas Data Repository.

Period of data retention

All the data and any publications will be made available as soon as they are created. The prototype BuDAS system will be made available when tests of stability have been completed to ensure it runs smoothly.

Data formats and dissemination

As mentioned previously, all the data will be made available on-line, via GitHub and the Texas Data Repository.

Data storage and preservation of access

The software will be stored on GitHub, which is a third-party open repository that allows software and data to be stored. We will provide links on our web pages to ensure search engines are able to discover and locate the repository for the project.

In addition, to ensure that the data will be discoverable by the larger research community, our data will be curated in the Texas Data Repository (<http://data.tdl.org/>) through the Texas Digital Library. The Texas Digital Library (TDL) is a consortium of academic libraries in Texas, providing shared technology services to support secure, reliable access to digital research collections. The Texas Data Repository is a project of the TDL to develop a statewide research data archive for researchers at Texas universities. Data will be curated in the repository following accepted standards. Data in the TDR are assigned DOIs, which

enables the citing of data by anyone who uses it. Metadata for the project data will reflect best practices developed by the Library of Congress (Metadata for Digital Content [MDC]) and will provide information on subject, provenance, authorship, methods and post-processing, and copyright that will support discoverability, curation, and preservation of the collection, as well as accessibility via harvesting and APIs.