

Data management plan

Digital files and assets to be created during the fulfillment and over the life span of the proposed VSim Project Repository and Archive include data related to the project in a general sense (grant proposals, specifications, progress reports, correspondence, etc.), the VSim software (the development environment, iterations of builds, etc.), and the web-based project repository and archive. Additionally, consideration must be given to the digital assets uploaded by outside contributors into the project repository. These assets will include both content generated in VSim formats (.vim, .nar, and .ere files) and multi-media content that is associated with one or more of the virtual environments and in standard file formats (.jpg, .pdf, .mov, etc.).

Roles and responsibilities

Co-Pis Snyder and Friedman will be responsible for the files and data generated during the project development phase. Snyder will be responsible for the work files related to the project generally. They will be stored on servers administered and maintained by the support unit of UCLA's Institute for Digital Research and Education, led by Grant Young. The IDRE servers are mirrored, routinely backed up, the software regularly updated, and the hardware regularly replaced (or kept within warranty and extended warranties) to ensure constant access. At the completion of the project, the work files will be encapsulated and given to the UCLA library for deep archiving. Friedman will be responsible for the files related to the development of the VSim software. As with development of the prototype, the programmers will use a locally hosted Git environment to control revisions and manage the source code. Once completed, the code will be pushed to GitHub for distribution to and expansion by the digital humanities community.

The UCLA Digital Library Program will have data management responsibility for both the web-based project repository and archive and all assets uploaded to the site by content creators and general users. This management responsibility will extend beyond the life of the grant period as the repository has been designed as a complement to the existing DLP offerings and as a solution to the challenges of archiving and providing access to academically generated 3D content.

Expected data

The activities of completing the proposed project will primarily generate files typical to the life cycle of any digital development (text files, XML spreadsheets, software code, etc.). It is only following fulfillment of the grant and the launch of the project that significant amounts of data will be produced. Content creators and users will be encouraged to share their 3D content and multi-media assets related to their virtual environments. These files will be uploaded to the repository and archive with appropriate metadata and organized for easy distribution. Content creators will be able to use VSim to generate three file types for distribution through the project repository and archive (.vsim, .nar, and .ere formats). The .vsim file is a binary aggregate built from the content creator's raw 3D model files (i.e., it is a single read-only file that includes all geometry, texture, and material information from the raw model files plus any information added by the content creator to control the display or dissemination of the environment). The .nar files include the information related to the display of linear narratives (e.g., the spatial coordinates and overlay information for each node). The .ere file format governs the display and launching of the embedded resources that can be associated within a virtual environment.

In addition to the 3D content and related multi-media files, the website will also maintain relational databases including, but not limited to, contributor accounts and tracking for the peer review process. The only data associated with the proposed project that might be considered private and/or sensitive is the user biographical information. For this reason, the user profiles and content management system will be hosted on a secure server to meet privacy standards.

Period of data retention

The archived project work files and the content uploaded the archive will be stored and preserved on the UCLA Library servers as an active repository until what time the content creators themselves request that their environments be removed or the archive is no longer viable in a costs/benefits analysis. The greatest challenge will be data storage. The amount of data generated for the project repository will grow exponentially as virtual environments and related multi-media content are uploaded by site users. Although the UCLA Library is committed to supporting and maintaining the archive as part of their hosted digital content, the need to cover the hard costs related to the site reinforces the need for the ongoing revenue stream envisioned as a result of the blended revenue strategy.

Data formats and dissemination

The web-based project repository and archive will use industry standard open source repository software (Fedora Commons) and metadata standards to ensure interoperability. New metadata standards will be written for the three VSim files types. These standards will be posted to the Fedora Commons metadata repository for reuse at other institutions should they choose to independently archive the content generated by their own faculty or students.

As the project repository and archive is been expressly created for the dissemination of VSim content and associated multi-media resources, the project team is committed to identifying strategies for simplifying access to and use of these materials. There are two scenarios where access may be restricted: if the content creators purposefully limit dissemination to a select group or if the content is only available to subscription holders per agreements with content-generating institutions or labs. (Even in such cases, content creators will be encouraged to provide some form of freely available content; e.g., a file that includes the nine phases of construction of a cathedral complex may only be available for subscription holders, but a file that includes only the current phase is free to the general public.) Content creators also have the choice between distributing read/write or read only copies of their narratives and embedded resource files.

Data storage and preservation of access

All UCLA Digital Library Program content is ingested into a Fedora Commons repository within the context of the Islandora digital library framework. The UCLA Library Islandora implementation at UCLA is distributed, with the repository hosted within a UCLA Information Technology Services data center, with replication at a University of California, Berkeley data center, in addition to conventional backup services. This level of replication and geographical distribution is afforded all content hosted by the UCLA Digital Library. In addition, the VSim project would leverage the services of the UC's Curation Center, including assigning permanent Archival Resource Keys (ARKs) and associated permanent URLs to all content through their EZID service, and utilization of the Merritt repository service for digital preservation (to be made available to non-UC VSim contributors at cost.) Any content that become a part of the UCLA Library's permanent digital collections would go into the Library's digital preservation workflow without additional cost to the repository.