

Proposed Research and Call for Participation:

How to File Place-Based Information Resources for Re-Use and Collaboration

Paul Cote, (your name here!)

Geographic Information Systems Specialist, Harvard University Graduate School of Design

December, 2007

Background: The primary activity of GSD students each term is a design studio. A studio is a project that considers design alternatives related to a specific place and proposed program of needs. There are approximately 50 studios each year, each having a similar research cycle that begins with intensive compilation of information about existing conditions; production of representations of many sorts intended to describe critical aspects of the existing condition; development of models intended for experimentation with alternative proposals; and finally, preparation of presentation documents. From the perspective of the majority of the Design faculty, the product of a studio is the collection of final presentation documents. However, it is also understood that the compilation of source materials and intermediate study models made by students in the course of the studio are also valuable after the studio is over. This proposal is oriented toward understanding how the source information and intermediate products of a studio might be organized in order to better serve the needs of a studio in progress, and also to facilitate the archiving and re-use of studio materials at a later date.

Although our terms and examples come from the world of Design Studios, the question of how the various information resources compiled for many research and development concerns are undoubtedly similar. For the purposes of maintaining a desirable level of generality we will refer to any of these compilation and generation activities as “Place-Based Projects” or Projects for short.

Proposed Audience: This research has two target audiences: The Research Assistant and Collaborator; and The Serious Archivist. To address the needs of the research assistants and collaborators, this research will provide a best-practices guideline, based on simple principles and tools, for organizing various types information resources in the project’s shared file space. This guideline is intended to be easy to follow in order that a large number of projects achieve easily understood benefits during the in-progress phase of the project with little extra effort. The second audience includes archivists and project leaders who are interested in the post-project phases of the information resources life-cycle. This audience will be interested in standards for documenting and formatting work for the purposes of archiving and cataloging and long-term re-use. Understanding and addressing these issues will require a deeper understanding of conventions and standards for documenting information resources. Working through both the ordinary and the serious approaches to this problem will help to assure that the simple guidelines are simply a simpler profile of the more serious approach.

Call for participation and comments: this one-page prospectus has been sent to people who may be interested and able to contribute to this project or who may have information about similar work that should be considered. At this stage, any advice related to the definition of the problem and scope; prior work in document management; metadata; archiving and cataloging; file format issues; or other issues will be most welcome.

Timeline: We hope to complete enough research to have a rough outline of Best Practices for the Research Assistant and Project Collaborator, prepared by January 1, 2008. This will include a first draft of a complete paper including the more in-depth concerns of the archivist.

Goals for Project Data Organization (business case)

1. A pre-existing convention for organizing project data that has been reviewed by a broad community of stakeholders will save project-planning time spent in inventing project-specific filing procedures, and save the trouble of learning hard data management lessons.
2. A well defined set of principles and associated technique for filing different types of information will reduce the amount of researcher time needed to figure out where a new piece of information should be filed, or where to look for a particular type or piece of information within the project file system.
3. A predictable directory structure and standard for documentation will produce economies of scale and other incentives for the developers of tools for harvesting metadata and creating cataloging and document management tools.
4. Project collaborators may develop analytic GIS models, maps or 3d models that integrate elements of base data or model elements developed by other project collaborators. A predictable and stable architecture for organizing this work will assure that these linked resources can be used on different filesystems and synced up with shared project resources integrated with a minimum of duplication or broken links.

Elements of a Project Shared File Space:

1. Photos, video files and other place-and-time specific observation data
2. GIS data and associated maps and analytic models
3. Statistical information and related data
4. 3d Models with associated component and texture resources, derived model products, e.g. STL files, etc.
5. Human-readable text or multimedia documents
6. Metadata: Provenance, Dates, keywords, georeferencing, License information

Dimensions of the Filing Problem:

1. Useful categories of information resources and their relationships. For example:
 - a. Information according to type and purpose e.g. site photos, GIS data and models, 3d Models and their resources
2. Useful categories of information with regard to their place in the project lifecycle. For example:
 - a. Original Source material: data from sources outside the project
 - b. Intermediate project data: models, maps or other products that may integrate some original source data with work accomplished in the progress of the project
 - c. Final project documents
3. Appropriate hierarchal directory or folder structures that provide a strong framework for inter-referencing between elements, documentation, collaboration, cataloging, discovery, and archiving.
4. Appropriate metadata standards for different types of data and different levels of investment in the archival enterprise. (simple readme files, Dublin core, marc, fgdc, ISO19115, ...)
5. Different activities associated with phases of the project lifecycle – from the data gathering phase, to the project collaboration phase, to the project completion and publication phase may indicate different levels of completeness and obsolescence for different pieces of the project filesystem.
6. Different means of cataloging, searching and archiving project resources: e.g. Harvard Geospatial Library , Harvard DataVerse Project, MIT's DSpace project, or cataloging and discovery means that can be applied using the active project filesystem.