

Library Lab Proposal: FACADE2: MIT and Harvard Collaboration

This resubmission of the original proposal addresses the long term thinking of the FACADE project, the understanding that we are submitting a proposal for a small piece of the whole project's development, our thoughts about long-term hosting and support, and the contributions from MIT and the Frances Loeb Library. New additions are noted with asterisks (*) within the following narrative.

*The Frances Loeb Library and the MIT Libraries are submitting a proposal that will allow a first step in the further development of a shared production tool, workflows, and a shared repository for the collecting, archiving, access and preservation of electronic architectural files. This proposal is focused on upgrading the current metadata tool for the tagging of architectural files, along with testing of the tool across two institutions. *

Changing technologies in the generation of architectural drawings have transformed architectural practice and that of other design disciplines, such as landscape architecture and city planning. As a result of these new technologies, architects and designers from across the globe can work collaboratively on a project around the clock. This has translated into an overwhelming number of born-digital records in a diversity of text-based and image-based files that, in some instances, should be captured and preserved in archival settings for teaching, learning, and scholarly research. Image-based file types currently range from raster image files (from scanning), to vector image files (from 2D CAD drawings), and 3D digital modeling files (from computer aided manufacturing technologies) produced by a diversity of evolving software products (characterized by technologies that continue to undergo rapid and continuous change). The capture and preservation of these born-digital records is indeed very much a concern both for architectural firms and architectural archives alike, and very few archives or architecture firms are able to address the issue.

At the Frances Loeb Library of the Graduate School of Design we have become increasingly aware, through conversations with faculty and alumni alike (all past or future potential donors of architectural records collections) of the complexity of dealing with born-digital materials that make heavy use of 2D drawings and/or 3D modeling. Although, in some instances, we have come to agreements as to general file format, general file set-up, compliance with layering and naming conventions in order for us to be able to receive some files, the conversation cannot really move forward until we can ensure the adequate preservation of, and future access to, these digital files. Archivists at many institutions and locales are dealing with the reality of CAD files that cannot be truly collected because best practices do not necessarily solve the dilemma of their accurate capture and actual preservation.

The capture, curation and preservation of these digital files has previously been studied, most significantly by two studies. One of them was an early study commissioned by the Art Institute of Chicago in 2004 (*Collecting, Archiving and Exhibiting Digital Design Data*), and the other was a study by MIT developed between 2006 and 2009 called *FACADE: Future-proofing Architectural Computer Aided Design* that created a collection of digital files from single projects provided by three contemporary architects who allowed them to access building project data to use as a research test bed. Most

groundbreaking in the FACADE study was the development of a “project information model”, an ontology related to process documentation (described in their report http://facade.mit.edu/files/FACADE_FinalReport.pdf as a “relationship between various CAD files and versions, and between CAD files and other project communication and documentation”) that needs to be further tested, developed, and refined. As part of the FACADE project, the research team also developed a tool referred to as the ‘curator’s workbench’ (CWB) that allows curators of architectural collections to add metadata through a web-based computer application. Through the CWB, metadata is added to the original filing system, as received, through tagging multiples files at one time and also allows the curator to identify significant ‘selected objects’ that could require additional archival description and preservation. In the project, the archived files and newly created metadata were saved to a network file system with the intent that in the future these files would be migrated to DSpace and possibly other archiving platforms. By the end of the MIT study, the approach proved to be viable from a *technical* point of view for the capture, curation, and preservation of CAD and other project files. Yet what emerged as real difficulties were access to the software via license keys, as well as intellectual property issues that arose for the architectural firms and their clients. As part of the project, a non-exclusive royalty free license was drafted, although it requires further review through legal counsel of architecture firms and academic institutions.

Given the strength of both Harvard and MIT in both the technical and/or curatorial aspects of the process of archiving architectural drawings, each institution has a vested interest in the strength of each other. Through the support of Harvard’s Library Lab, we would have the possibility of collaborating, through this joint study (FACADE2: MIT and Harvard Collaboration) towards what we both envision in the long run as an open source platform with shared infrastructure for the effective capture and preservation of architecture drawings. The focus of this collaborative project work would allow Harvard and MIT to further refine the CWB, then to critically test the system using new data from donors that Harvard and MIT would provide (each institution will provide at least one design project). The project would also allow us to further refine the ontological model, and would allow both institutions to work collaboratively in reaching agreements in relation to institutional IP issues and in relation to how to approach the legal concerns of donors. We currently view the rights management issue as an active process to resolve; and understand that, as a parallel process, we will have to allow for the embargo of some materials as different categories of information are analyzed. In particular, we will want to investigate the option of determined embargo periods with donors for the parts of collections we will work with that contain sensitive material or that firms are legally mandated to retain. Furthermore, the ultimate result is that we could move the collaboration forward toward a shared production tool for the capture and preservation of and access to born-digital CAD files for teaching, learning and research.

Following up on the work done by the two previous research studies as a collaborative effort between Harvard and MIT will help deal with one of the most difficult archival problems we have as librarians/archivists – the preservation of architectural documentation. The solution for this “cultural problem” cannot be taken on by one institution alone, but as a collaborative effort we can take the next steps to move from a research prototype to a production archive that can be used in archival practice

for both MIT and Harvard. Further, the open source nature of the work will enable other institutions to find a path to managing architectural archives of the 21st century. The use of DSpace at MIT, a well-used repository in many institutions (including Harvard) will allow us to incorporate this new material into our current archival workflows without building a new repository. We believe this project proposal leads in the direction of supporting the capture and preservation of architectural drawings, and as such will greatly enhance the documentation of architectural practice and the mission of architectural archives at both the MIT Library and Harvard, and contribute to that effort more broadly.

Needless to say we will share information with the project team for Zone 1, another joint Harvard – MIT project currently funded by the Library Lab for the development of a rescue repository for faculty papers and keep abreast of the developments in relation to the rescue repository. We are aware that Harvard's Property Information Resource Center (PIRC) and MIT's Department of Facilities, offer potential partners with whom we can work to obtain feedback and we expect to engage them in our thinking in this project.

Scope of Technical Work:

The 'curator's workbench' (CWB) is a web-based tool for creating, editing and managing metadata for an architectural project, specifically the metadata pertaining to individual project files that constitute the corpus being described. The metadata conforms to a specific ontology known as the 'PIM' (Project Information Model – loosely derived from an emerging building industry standard known as 'BIM' – Building Information Model). Thus, the CWB is a PIM instance editor. Numerous other tools, scripts and workflows, and procedures also contribute to the generation of a complete PIM instance, but the CWB represents the primary curatorial and administrative interface to it. The CWB underwent substantial iterative development in the FACADE research project, as various needs and requirements (e.g. refinements or expansions of the PIM itself) emerged, but little effort was made to 'harden' it into a production-ready tool. The work this proposal describes would address this need in large part, as well as 'rationalize' or streamline the workflows surrounding the CWB, to facilitate its use as the primary tool to describe an architectural project. The proposed work falls into two categories: infrastructural and functional.

Infrastructural deliverables include a complete review of the software 'stack' – i.e. the components that are assembled to make up the application. Since the CWB was written in a research context, components like the RDF 'triplestore' (the semantic 'database' in which the PIM resides) were often not fully supported or experimental. Other components are really just out of date, and need to be upgraded or replaced. Also in scope is the addition of industrial-strength application logging and instrumentation in the CWB code, so that it can be maintained and bugs more easily discovered and fixed. The primary functional deliverables fall into three areas:

Initial PIM generation/definition. The process of defining the basic project data – essentially formal namespaces, etc currently falls outside the CWB, and requires numerous error-prone manual editing

steps. We would build this functionality into the CWB itself, perhaps as a 'new project' wizard, the better to manage and validate this data.

Derivatives. An important part of curating collections is the creation of so-called 'derivative files' – typically produced by specialized CAD software with the aim to ensure preservation and friendly access. These become 'special objects' in the PIM and the process of integrating them into the existing project again requires hand-editing outside the CWB. We would integrate this fully into the CWB.

The CWB contained partial and experimental support for expressing and attaching rights metadata to project files in the PIM instance. We would finish this work, and provide any additional query, tagging or other functionality associated with rights assignment.

The success of this project will be evident in the streamlined processes using the Curator's Workbench, and in improved workflow from the time of data acceptance from a firm through the curation process and through to delivery of information.

We believe that funding from the Library Lab will help to move the concept of the FACADE project from research to production, allowing both MIT and Harvard to begin the process of accepting collections that contain CAD and other files because we will be able to preserve and make them accessible for teaching, learning, and research.

*In the original FACADE grant awarded to MIT, the assumption was that other institutions would benefit from MIT's development of an open source tool and the use of the open source DSpace repository functionality. MIT found during the grant work, however, that many institutions are not yet able to take on support of this infrastructure themselves. This Library Lab proposal will further define the development roadmap we will follow to bring FAÇADE to a production level service.

*The contributions from both Harvard and MIT will include:
Technical support from the Berkman Center technical team

Two members of the MIT Libraries Software and Development (SDA) team will work with the Berkman Center team to provide an initial 'discovery and knowledge transfer' for the CWB codebase in which SDA staff transfers copies of CWB code and documentation to the development group. There will doubtless be several questions around establishing a mutually agreeable development environment, etc. We expect this phase will consume 20-30 hours of staff time, minimum.

This will be followed by a 'consultative' phase as the development work takes shape. There will be brief weekly status meetings (via phone/skype, etc), as well as asynchronous communications (email) as issues arise. This will take 4-5 hours of SDA staff time on average for the duration of the update/development cycles, which are likely to comprise maybe 6 months of the total project.

This yields a project total of 116-150 hours.

Time from curators at both MIT and Harvard will be approximately 80 hours per architectural project each, assuming that each institution works on testing one project in the upgraded tool. This is time

