Curating Architectural 3D CAD Models

MacKenzie Smith,

Associate Director for Technology,

MIT Libraries

July 2008

Abstract

Increasing demand to manage and preserve 3-dimensional models for a variety of physical phenomena (e.g., building and engineering designs, computer games, or scientific visualizations) is creating new challenges for digital archives. Preserving 3D models requires identifying technical formats for the models that can be maintained over time, and the available formats offer different advantages and disadvantages depending on the intended future uses of the models. Additionally, the metadata required to manage 3D models is not yet standardized, and getting intellectual proposal rights for digital models is uncharted territory. The FACADE Project at MIT is investigating these challenges in the architecture, engineering and construction (AEC) industry and has developed recommendations and systems to support digital archives in dealing with digital 3D models and related data. These results can also be generalized to other domains doing 3D modeling.

Computer-Aided Design in Architecture

Use of 3-dimensional Computer-Aided Design (3D CAD) software is routine in most design professions, including architecture, engineering, archeology, conservation, and other disciplines that deal with the built world – past, present, or future. 3D modeling is also routinely used in science, technology and medicine to describe and visualize physical phenomena. 3D is increasingly used in interactive teaching tools and virtual environments like Second Life[1]. 3D models describe the geometry and other attributes of physical objects, and are dynamically visualized with computer graphics. Compared with most other digital data formats, 3D models are very complex and the formats used to store them are typically proprietary to a particular software product, presenting further challenges to long-term curation. Despite the prevalence of 3D models among the data that digital archives are ingesting and the many challenges they present for long-term curation and preservation, very little research has been done on curating them.[2]

In the field of architecture, 3D CAD software is becoming the design tool of choice. There are many software vendors supplying the architecture, engineering and construction (AEC) industry, most notably AutoCAD and Revit from Autodesk[3], CATIA from Dassault Systèmes[4], and Bentley's[5] Microstation and GenerativeComponents. 3D models are increasingly employed across all phases of building projects, and are beginning to be seen as useful for ongoing product lifecycle management (e.g., for ongoing building or aircraft maintenance).

The workflows that use 3D CAD in building projects are currently ad hoc and prone to translation errors between the varying products deployed by different parties involved in a project. This has led the AEC industry to define a "Building Information Model"[6] as a standard way to model buildings in 3D. BIMs include the 3D CAD model plus significant properties of the model (e.g., features and materials) to facilitate communication between the different parties involved in the project and the future building owners. A few CAD modeling products claim support for BIM (e.g., Revit, GenerativeComponents) but the standard is still evolving so adoption by the industry is limited. In architecture projects, as with most projects, the 3D CAD models are just one of many types of files produced that are important for curating the collection over time. From the 3D model many more specific 2-dimensional drawings are produced for particular aspects of the building (electrical, HVAC[7], floor layouts, etc.). There are also artistic renderings produced for clients – photorealistic images, photographs and

---

[1] Second Life http://secondlife.com/
[2] Two major projects exploring 3D models from the engineering sector are the "Immortal Information and Through-Life Knowledge Management (IITKM): Strategies and Tools for the Emerging Product-Service Paradigm UK Knowledge" (https://www-edc.eng.cam.ac.uk/kim/ and the Digital Engineering Archives project directed by William Regli of Drexel University http://gicl.cs.drexel.edu/wiki/Main_Page#Active_Research_Projects
[3] Autodesk's website http://autodesk.com
[4] Dassault Systèmes website http://www.3ds.com/
[5] Bentley Systems website http://bentley.com/
[6] The Building Information Model (BIM) standard is being developed by the National BIM Standard organization, described at http://www.facilityinformationcouncil.org/bim/index.php
[7] Heating, ventilation and air conditioning

videos of the construction site – and voluminous email and document exchange between architects, clients, contractors and other parties (e.g., all the official Requests For Information and corresponding Architect's Supplemental Instructions).

Figure 1. The MIT Stata Center, designed by Frank O. Gehry, completed in 2004 and designed using Dassault Systèmes' CATIA.

While the 3D models are of significant interest, the contextual files are often necessary to interpret them correctly. For example, client presentations help to explain an architect's design intent as implemented by a model, so researchers looking at the model need the context of the presentation to make sense of it. This illustrates the need to correctly define the designated community and target audience for collections including 3D models, since curation strategies could differ significantly for different designated communities.

3D CAD Preservation Challenges

In the field of digital preservation, 2D drawings produced by products like AutoCAD present few new problems since they are industry standards (i.e., the DWG and DXF formats) and they can easily be converted into other standard 2D formats such as JPG or PDF. Neither are 2D drawings significantly interactive in their native systems, so requirements for their long-term preservation are similar to other static formats like visual images. For situations in which 2D CAD is the best representation of the object, standard formats and migration preservation strategies can reasonably be applied.

However 3D CAD is a different story. 3D models are created in proprietary software using non-standard native formats, and each product uses different techniques for capturing a model's shape information via designer-specified parameters, storing geometry and other properties attached to the geometry, and rendering the model on

the computer screen. Each software product's method of doing this uses different, very complex mathematical techniques, for example parametric B-Spline or NURBS equations, non-parametric equations, or a combination of both (Lee, 1999). These different methods are the differentiators that define competitive advantages among such products; there are few incentives in the industry to define standard parametric data formats.

Because of this, native 3D CAD file formats cannot be interpreted accurately in any but the original version of the original software product used to create the model. There are no products that support export of internal 3D parametric models into a standard, neutral format, nor are there standard formats that support parametric data. One CAD encoding standard from the engineering industry – Standard for the Exchange of Product (STEP) – is adding support for parametric models, but that has not led to product support so far. This translation problem even extends to different versions of a single product, with notable translation errors occurring[8].

There are standards for data export between 3D CAD products but they all incur information loss from the original native parametric model since they reduce the model to static geometry and attributes – solid, surface, or wireframe – none of which include the architect-specified parameters that signal design intent. Given that limitation, most of these standards are mature and widely supported. They include IGES[9], STEP[10], IFC[11], STL[12] and the emerging PDF/E[13] ISO standard from Adobe.

For the foreseeable future, in order to continue interacting with the native 3D CAD parametric model, it will be necessary to maintain the original, proprietary software that created it. This situation is similar to that of custom computer games, virtual worlds, and many other software applications that are used in teaching and research today. A requirement to maintain proprietary software over time inevitably leads an archive to provide support for software emulation, but emulation of arbitrary software products is still an extremely complex and expensive operation that is largely untried in production environments.

---

[8] An October, 2005 article in BusinessWeek described problems incurred by the Airbus company developing its new A380 airplane when they upgraded from version 4 to version 5 of CATIA at different times across the design and manufacturing groups, leading to design incompatibilities and huge financial losses. http://www.businessweek.com/globalbiz/content/oct2006/gb20061005_846432.htm

[9] IGES (Initial Graphics Exchange Specification) was initially defined in 1979 by a group of CAD vendors and customers with support from the National Bureau of Standards (now NIST) and the US Department of Defense as NBSIR 80-1978. It is still widely deployed but deprecated in favor of STEP.

[10] STEP (Standard for the Exchange of Product) model data, is ISO standard 10303 and covers computer-interpretable representation of product data throughout its lifecycle. It consists of hundred of "parts" describing different aspects of the data, and "application protocols" for different types of products or industries. For more information, see http://www.tc184-sc4.org/SC4_Open/SC4%20Legacy%20Products%20(2001-08)/STEP_(10303)/

[11] The Industry Foundation Classes (IFC) specification was developed by the International Alliance for Interoperability (IAI) for the AEC industry, and is described on the IAI website http://www.ifcwiki.org/

[12] The STL (stereolithography) format was developed in the 1980s by 3D Systems, a company that makes rapid prototyping machines for built objects (i.e., products) and is widely supported in the CAD software industry for surface geometry

[13] PDF/E is described on the Adobe website http://www.adobe.com/products/acrobat3d/

Preservation Strategies for 3D CAD

The MIT Libraries are conducting the FACADE Project to investigate curation and preservation strategies for digital designs and related data from major architecture projects that include 3D CAD. FACADE – Future-proofing Architectural Computer-Aided DEsign – is working with several well-known architectural firms including Frank O. Gehry, Moshe Safide Associates, and Morphosis (i.e. Thom Mayne) to understand their data production workflows and subsequent data use, and their expectations as a designated community for digital archives of these projects.

3D Format Identification

The MIT Libraries and Archives use the DSpace platform[14] to archive scholarly research in digital formats, so the project has implemented its strategies on that platform as a proof of concept. The DSpace platform currently supports minimal preservation of all digital formats (i.e. bit-level preservation) and functional preservation for selected formats by standard techniques such as format migration. The level of preservation intent is signified by local policy defined in an internal format registry, and the curating organization is responsible for ensuring that supported formats have viable preservation strategies defined, and that these are implemented and maintained over time. As part of the FACADE Project, MIT has made several major changes to digital preservation support in DSpace. They include fine-grained format identification, by integrating DSpace with the format identification tools DROID[15] and JHOVE[16], and with the external format registries PRONOM[17] and GDFR[18], so that preservation strategies can rely on third-party registries to help monitor the status of 3D formats and identify migration tools as necessary. We are also acquiring format information for both proprietary and open standard CAD formats into the format registries for future reference.

3D Format Migration

Data acquired by the project was exclusively in proprietary native formats produced by CAD software tools used for the project. Given the difficulty of working with those formats over time, we have evaluated the available standard 3D formats, and are recommended a format migration strategy to minimize the risk of permanent data loss. For each 3D file we ingest, we create a standards-based version in either IFC or STEP[19] by exporting the file from the original software in the standard format. This is a manual process requiring expertise in both the native software and its underlying data model (e.g the CAD model tree) to create useful standard versions. In future this export process may become more automated, but in the meantime it requires archives staff to have some expertise in CAD software and probably its normal use by the domain of interest.

---

[14] DSpace is detailed on the project website at http://dspace.org
[15] DROID http://droid.sourceforge.net/wiki/index.php/Introduction
[16] JHOVE http://hul.harvard.edu/jhove/
[17] PRONOM http://www.nationalarchives.gov.uk/pronom/
[18] The Global Digital Format Registry (GDFR) http://hul.harvard.edu/gdfr/ will be released in August, 2008 so discussion of it here is with respect to the specification and other documentation of the registry's design and data model. The external registry support in DSpace is abstracted to allow for multiple and changing external registries.
[19] Some CAD products export to STEP but not IFC, for example, CATIA, while others support IFC but not STEP (e.g., Revit) so that no one standard format is practical in all cases.

In addition to the standard STEP or IFC format, which captures the solid geometry of the original 3D model, we recommend creating another, simpler standard format – IGES – to capture the surface geometry and parametrics of the original model. This produces a less functional version than STEP or IFC, but is less prone to translation errors during export. Finally, we recommend creating a presentation version that is readily accessible on the Web – 3D PDF – with the understanding that this version will probably need to be replaced fairly often as Web formats for 3D evolve. With these four versions – original, standard, geometry, and presentation – we feel that the needs of our designated communities can be reasonably met.

3D Format Emulation

The FACADE Project has also explored the implementation of a software emulation framework for DSpace to support native CAD files requiring original software to interpret. Current virtualization and paravirtualization platforms (e.g., VMWare[20], QEMU[21], XEN[22]) were tested to render older 3D models. While it proved possible to run several relevant CAD products (e.g., CATIA version 4) in a virtualization environment, these products' licensing protection mechanisms (e.g., requiring a valid license key) appear to be an insurmountable difficulty with this strategy. Additionally, while virtualization systems are easy to acquire and will work with these products, to achieve very long-term preservation we anticipate the need to use true emulation products rather than those requiring a particular chipset to virtualize the target platform. For these reasons, while we recommend retaining the original native CAD format in the archive, we believe there will be significant challenges in working with those files in the very near future since they require the original software product to open.

Project Information Models

One important finding of the FACADE Project is that, at least for large architecture projects, a 3D model is of most value to a designated community (e.g., future researchers, historians, design professionals) if it is available in some context that helps to explain the design intent it implements, and any problems that arose from the design during construction or use of the physical artifact. So we are ingesting collections of data that are structured around the 3D model but relate it to the other data from the project into a "building project" collection. We call this structure the "Project Information Model" (PIM) to distinguish it from the "Building Information Model" described earlier. A BIM can capture the 3D model and many related properties, but it does not capture the voluminous correspondence and presentations that the building project generates. So the PIM structure that we have designed includes 3D models (or BIMs, if they are provided) as well as the 2D drawings and all the other material in the collection. We have designed the PIM as an RDF ontology, and created test data manually for our initial prototype while we developed software to assist non-programmers in creating PIM metadata for subsequent projects.

The PIM ontology currently supports the notion of a "file" that is assiged a set of metadata properties of general use (i.e., building phase, zone, architectural discipline,

---

[20] EMC's VMWare product is described on the company's website http://www.vmware.com/

[21] QEMU is an open source software machine emulator and virtualizer http://fabrice.bellard.free.fr/qemu/

[22] XEN is another open source software project originally developed at the University of Cambridge and now commercially supported by XenSource, Inc. http://www.xensource.com/

file format, file type). It further defines a "design object" (e.g., 3D model or 2D drawing set) and a "presentation object" (e.g., a client presentation or other file with significant relevance in defining design intent). Together these properties provide the data we need to build the public User Interface to the collection, via search and faceted browsing, and there are additional properties used for digital curation and preservation.

Once we have created PIM metadata – an RDF graph representing all the data we have collected and created, including all the 3D model derivatives – we can add that file to the collection archive for the building, and pass it to external discovery and navigation tools that support user exploration of the collection. Users search, browse and navigate the collection, and retrieve data files of interest from the archive into their browser. Our recommendations insure that there is at least one Web-renderable version of every file, including the 3D models (usually 3D PDF) so that users are not required to have the native CAD software loaded on their local machine just to explore the collection. However we also recommend retaining the original 3D format so that users who do have that software, and want to interact with the model, can do so.

The PIM ontology has evolved considerably during the project and we have published the current version, along with data from our test projects, on the FACADE wiki. We are also developing prototype software to assist curators with creating PIM metadata (the "Curators' Workbench") and will release that as open source software at the project's conclusion.
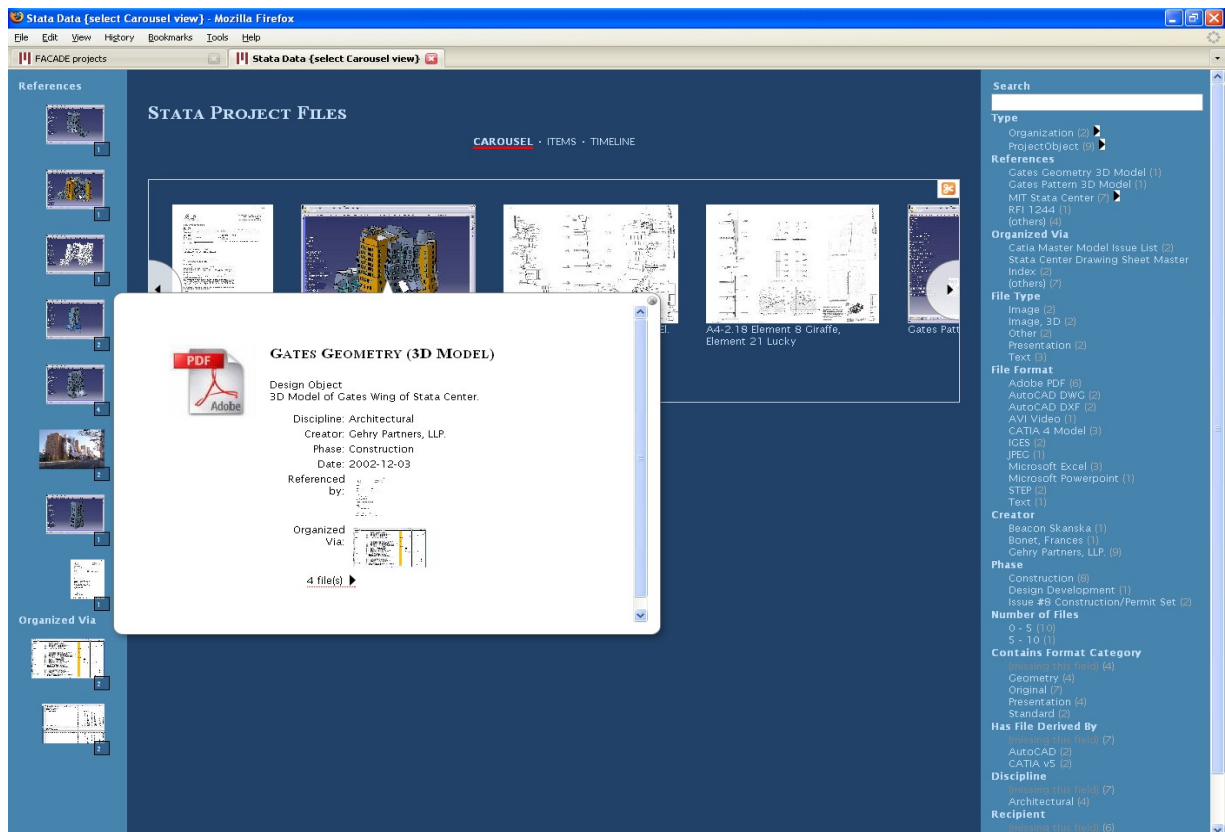

Figure 2. Screenshot of the first prototype FACADE public User Interface for the MIT Stata Center

Finally, we have developed a prototype end-to-end system that allows users to discover, search, explore, and visualize the data in our test collections, and that code is now available as free, open source software for other organizations to deploy locally[23].

Conclusions and Future Work

3D models are rapidly becoming a critical part of the scholarly record, but these models and their visualizations present many new challenges for digital archiving and preservation, some of which include:

- Defining the target audience for the 3D models, their visualizations and related data. Does the designated community consist of future students, researchers, historians, practitioners, records managers, the general public, or those responsible for managing the original physical artifact (e.g., a building or body part)?
- What functionality is expected of 3D models? Full interactivity or just a static visualization of the model? If interactivity is needed, is it acceptable to re-parameterize the model when the need arises, or must it be in that state at all times? In other words, how important is the authenticity of the model (i.e. the original parametric model)?
- Is the object of preservation the model itself, or the design intent and process that it represents? Is related documentation for the design-and-build process necessary to archive and preserve together with the model?
- If original parametric 3D models cannot be saved, is static, solid geometry sufficient? Or simply surface geometry or wireframe versions? What about just documents that visualize the model in TIFF or PDF?

Once the designated community or target audience is defined and their initial requirements for use and reuse of 3D models established, the decisions on what data to capture, what additional versions of that data to create, and how to preserve the data become possible. Choosing among standard formats for encoding models is possible so that best practice can be identified for a given designated community. Moreover, selecting what additional project data should be archived together with the 3D model (e.g., 2D drawings, images and videos, correspondence and formal project documentation) also becomes easier. The work then consists primarily in creating the derivative versions for preservation, and designing the structural metadata (relationship map) for a collection so that the relationships between its parts are clear and can be rendered on screen for students, researchers, practitioners and other interested parties.

The FACADE Project has developed such a set of recommendations for best practices for architecture projects that captures a variety of interesting data, relates them together for later use by several designated communities, and provides preservation strategies for the 3D models. We believe this work can be extended to additional 3D modeling products and into new domains of practice that use 3D.

---

[23] The software to search, browse and navigate the architecture PIMs is based on work from the MIT SIMILE Project http://simile.mit.edu/. It is already integrated with the DSpace archive platform so that the data can reside in a preservation environment but user interaction with the collection happens outside the archive. New software supports bulk import into DSpace of all the project files and related RDF PIM data. At the project's conclusion all the software for the prototype system is being made available via the project's website at http://facade.mit.edu

An issue still to be resolved is the appropriate intellectual property rights for architecture collections. Since the material is all digital, standard gift agreements used by archives are not entirely appropriate – there is no need for exclusive copyright transfer to the archive, a royalty-free, non-exclusive license to archive, preserve, and disseminate the collection is sufficient. But since these collections often include highly sensitive business records there is an understandable reluctance by the architects to allow the material to be made publicly available. In the past, embargoing these materials for decades would have been acceptable practice, but given the effort involved in accessioning and ingesting digital collections (and preserving them for even a decade) the inability to disseminate the collection is a concern for libraries and archives. We are exploring licensing options with architects and their professional organizations (e.g., the American Institute of Architects) to determine what middle ground is acceptable to both creator and archive for the long term.

Acknowledgements

References

Lee, K. (1999). *Principles of CAD/CAM/CAE systems*. Reading, MA: Addison, Wesley.