

MacKenzie Smith

Associate Director for Technology, MIT Libraries
Directrice déléguée à la technologie pour les bibliothèques
du Massachusetts Institute of Technology (MIT)

Abstract. The FACADE project at the Massachusetts Institute of Technology (MIT) in the USA is exploring the long-term archiving and preservation of digital data from architectural projects in general, and 3D models and related information in particular. The project is working with design data from major building projects of the architects Frank Gehry and Moshe Safdie initially, and will generalize its findings to other architectural firms using 3D design software. The project is also exploring how to relate 3D designs with related 2D drawings, digital images and videos, email and other communications, and with emerging Building Information Models (BIMs). A prototype archive for this material will be produced with DSpace, the open source digital archive system originally developed by MIT and HP Labs and now widely used by research organizations world-wide.

KEYWORDS: Digital archives, digital preservation, 3D CAD models, building information models, digital architectural records, research.

Résumé. Le projet FACADE du Massachusetts Institute of Technology (MIT) aux États-Unis s'intéresse à l'archivage à long terme ainsi qu'à la pérennisation des données numériques issues de projets architecturaux. Il cible en particulier les maquettes en 3D et les informations afférentes. Dans un premier temps, nous travaillons à partir des données issues des grands projets des architectes Frank Gehry et Moshe Safdie, en vue d'une diffusion auprès des agences d'architecture utilisant des logiciels de conception en 3D. Le projet FACADE cherche également à associer les maquettes 3D aux autres fichiers et données d'un projet (plans 2D, images numériques, vidéos, courriers électroniques et autres supports de communication) ainsi qu'aux Modèles d'information du bâtiment (BIM). Un prototype d'archivage de ces éléments sera produit par le logiciel DSpace, la plateforme d'archivage numérique (logiciel libre), conçue à l'origine par le MIT et HP Labs, qui est aujourd'hui utilisé par de très nombreux centres de recherche dans le monde entier.

Mots clés : Archives numériques, conservation numérique, maquettes virtuelles 3D, Modèles d'information du bâtiment, archives numériques d'architecture, recherche.

Session 3: Long-Term Preservation of Digital Archives
Session 3 : La pérennisation des archives numériques

FUTURE-PROOFING ARCHITECTURAL COMPUTER-AIDED DESIGN: MIT'S FACADE PROJECT

Computer-Aided Design in Architecture

The use of modern 3D CAD software has become routine in many architecture and design firms, as well as for building restoration and preservation projects. 3D software is also increasingly used for interactive teaching in architecture schools, and for virtual built environments like Second Life.¹ 3D models describe the geometry and other properties of physical objects, and are dynamically visualized with computer graphics. While older, established architects often still rely on paper drawings and physical models, younger design professionals are working entirely by computer and their vernacular language is 3D CAD. A few major international architects – Frank Gehry, Moshe Safdie, Zaha Hadid, and others – have been using complex 3D modeling software for years in the design and construction of their buildings.

As one example, the MIT Stata Center was designed in the late 1990s by Gehry Partners, LLP (Frank Gehry's architectural firm)² using CATIA version 4 software, and the collection of digital files from the project numbers in the hundreds of thousands and requires 140Tb of disk storage. Three years after the building's opening in 2004 the original CATIA 3D models are already obsolete, and the institutional memory of the project's evolution is rapidly disappearing. The as-built printed 2D drawings have been saved, but given the complex geometry of the building they are inadequate to really understand the building. If the data from this project is not brought under archival control very soon the entire collection will likely become unfindable and unusable, less than ten years after the design process began.

There are many CAD software vendors supplying the architecture, engineering and construction (AEC) industry today. Popular software products include AutoCAD and Revit from Autodesk,³ CATIA and SolidWorks from Dassault Systèmes,⁴ Microstation from Bentley Systems,⁵ and Parametric Technology Corporation's Pro/Engineer.⁶ 3D models are increasingly used across the design, engineering and construction phases of building projects, and are beginning to be used for the ongoing product lifecycle management (for example, building maintenance). The mechanisms for using 3D CAD in building projects is currently ad hoc and prone to translation errors so the architecture, engineering and construction (AEC) industry is developing a standard "Building Information Model"⁷ to define a common way to model a building in 3D with all features and other information necessary to communicate between parties involved in a building project, and the future building owners. There are emergent products claiming



MacKenzie Smith

23

support for BIMs, but the standard is not yet complete and exactly what it will include or how open it will be is still unclear.

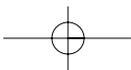
In architecture projects, there are many types of digital files produced during design and construction that are important for long-term archiving. In addition to the 3D CAD models there are hundreds or thousands of detailed 2D layer drawings produced for particular aspects of the building (Heating, Ventilation and Air-Conditioning – HVAC – and electrical systems, floor layouts, etc.). There are project “outputs” – for example, drawings or sketches of the building presented to clients at important stages of the project.⁸ There are photographs and videos of the construction site, websites about the building, and other multimedia related to the project. There are copious communications between architects, clients, contractors and other parties, including email, RFIs (requests for information) and ASIs (architect’s supplemental instructions). To represent a building project completely all of this digital information should be captured and linked together into a collection that can be easily searched and navigated, and preserved over time. The FACADE Project refers to these collections as “Project Information Models” or PIMs.

As we discuss these issues with architectural firms using digital technology, they clearly feel growing pressure to manage all of their own digital data so that it is findable and re-usable for their own business needs. Questions of why certain things were done, when, by whom, and so on are as difficult for the architectural firms to answer now as they will be for historians in one hundred years.

There are also a large number of projects that use 3D and related digital technology to help understand, manage, restore and preserve existing buildings. Particularly in Europe, a large number of historic building conservation and preservation projects are using laser scanning technology to create point clouds and 3D parametric models of buildings as a way to plan major renovations. The same ability to do structural and other analyses with a 3D model that modern architects value is also increasingly useful for all older buildings as well. These heritage projects are generating a lot of digital design data that will need to be managed and preserved alongside the data from new buildings not yet constructed.

Preserving 3D Models

Most digital data generated by architecture projects is straightforward to preserve using well-understood techniques from the digital preservation domain. For example, 2D CAD drawings such as those generated by AutoCAD software can be stored using industry standard formats (e.g. DWG or DXF) and converted into international standard formats such as PDF/A for long-term archiving. These PDF/A drawing files can then be monitored and migrated to newer formats as necessary. Tools and systems to do this already exist so that for simple 2D CAD drawings standard preservation strategies can reasonably be applied. For other building project data – images,



Future-Proofing Architectural Computer-Aided Design: MIT's FACADE Project

videos, correspondence and other communication data – there are similar preservation techniques that can be applied today.⁹

However preserving 3D CAD models over time is more complicated. 3D models are created with proprietary software using “native” data formats, and each software product has a different technique for representing the model’s geometric shape information and related properties, and the user-supplied parameters (the parametric model). These techniques involve very complex mathematical equations (e.g. parametric B-Spline or NURBS equations, non-parametric equations, or a mix of both)¹⁰ that differentiate each product and define its competitive advantages. Because of this, there are few incentives in the CAD software industry to define standard parametric data formats for internal data representation, and native 3D CAD models cannot be accurately interpreted by any software except the original version of the original software product used to create it. This problem extends even to different versions of a single product, with notable translation errors occurring.¹¹ With the emergence of industry standard Building Information Models (BIMs) this situation may improve since it will be necessary for many partners in the AEC building process to communicate via the 3D CAD model, but it is not clear when BIMs will become standard practice in architecture.

In the future, in order to interact with a native 3D CAD parametric model, it will be necessary to maintain the original, proprietary software that created it, as is true for other interactive software such as computer games and virtual reality programs. If the native 3D CAD models are kept, then the proprietary software used to create it must also be archived. Software is very difficult to preserve over time, requiring sophisticated operating system “emulation” in order to continue using it. Emulating arbitrary hardware and operating systems is a complex and expensive operation, and would be difficult to implement at every archive that has digital architectural records to maintain.

There are international 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 – solid, surface, or wireframe – and attributes, and do not include the user-specified parameters that informed the design intent. Given this limitation, most of the 3D data standards are mature and widely supported. They include IGES,¹² STEP,¹³ IFC,¹⁴ STL¹⁵ and the emerging PDF/E¹⁶ standard from Adobe. FACADE is initially focused on the IFC standard as our preferred archival master format, but we are also investigating the STEP standard since it is widely used in other industries and likely to be required by the U.S. Federal Government for their building design documents.

Archiving Digital Architectural Records in DSpace

As part of the FACADE Project, the MIT Libraries will create a prototype archive of digital architectural data in DSpace, an open source platform for archiving digital

MacKenzie Smith

scholarly research.¹⁷ The DSpace platform supports minimal preservation of all digital formats (i.e. bit-level preservation) and functional preservation of some popular formats using standard format migration techniques. The level of preservation support for a particular file format is specified by local policy via an internal file format registry, and the curating organization is responsible for insuring that “supported” formats have preservation strategies defined and implemented for their archive.

To be able to archive digital architectural data, FACADE is improving DSpace’s support for file format identification and validation, and its support for format description (i.e. technical metadata for file formats). This work is building on existing tools like DROID¹⁸ and JHOVE¹⁹ for format identification, and DSpace will also be integrated with international format registries (PRONOM²⁰ now, and GDFR²¹ in the future) so that preservation strategies implemented in DSpace can take advantage of third-party format registries to monitor the status of archived formats and identify migration tools in the future.

FACADE will also design and build packagers for Project Information Models that include the 3D and 2D CAD files, the multimedia and communication files. A METS profile²² or similar structural packaging standard will be defined for this purpose, and ingest tools created for the DSpace platform to support the curation, search and navigation of these collections via the Web. It is unlikely that these packagers will work for the data provided by any architecture project in particular – all of them will need manual intervention to supply information like the firm’s file naming conventions for its 2D layer drawings. But we believe that a toolkit approach may prove useful to archives that receive large project collections and want to ingest them into their digital archive.

To aid searching and identification of collected digital archives, the project will create simple collection-level (or PIM-level) descriptions in Dublin Core. More detailed metadata will be also created for a small number of significant files (e.g. the 3D CAD models, major project output files, or key communication files). In addition to descriptive metadata, the project will capture as much technical metadata about the digital files as possible during the ingest process (automatically, wherever possible) and will store this information in DSpace for ongoing curation. Finally, structural metadata about the entire collection will be created so that users may navigate through the collection in a graphical user interface via links. This will initially be done with a RDF ontology²³ so that the relationships between the parts of the PIM can be of any appropriate type.

A final area of work that FACADE will explore is a simple emulation framework for DSpace. This is to test the ability of a digital archive to run older software in a different operating environment. For example, running CATIA version 4 (the CAD software used by Frank Gehry’s firm for the MIT Stata Center project) in DSpace on the UNIX operating system. From this experiment we will learn what is possible now using off-the-shelf virtualization and emulation software (e.g. VMWare²⁴ or QEMU²⁵)

Future-Proofing Architectural Computer-Aided Design: MIT's FACADE Project

but we are not optimistic that this will lead to immediately useful digital archiving operations for the reasons given earlier.

Conclusions

3D models have become an important part of the scholarly record of modern architecture and design, as well as historical research and reconstruction, but these models and their visualizations present many difficulties for digital archives. Most digital archives today avoid them and ingest only files in simpler data formats (documents, images, etc.) thus an important part of the historical record – part of the architect's original design and intent – risks being lost. And as Building Information Models become more common in large architecture projects, the reliance on 3D CAD models will increase. Understanding these architecture projects in the future will require having access to all the data about the project from design through construction, including all the files about the building process itself.

Another important use of digital architectural archives will be teaching. Architecture and design students today do not have access to a library of digital design documents. They learn the tools and techniques of digital design, but without the benefit of exemplars from the field of practice. There is no equivalent of a slide library of digital designs. They also lack access to information about modern building processes as are embodied in Building Information Models and our planned Project Information Models. As one outcome of FACADE, we will develop scenarios for using digital designs in teaching so that architecture and design schools can see the benefit derived from having access to such archives.

As a practical matter, the “designated community”²⁶ or target audience of FACADE is the architecture and design professionals who use CAD in their daily work. Researchers and instructors are also intended beneficiaries of these digital archives, but since digital archives are not common today it is difficult to get feedback from researchers and instructors on what they need. Practitioners, on the other hand, recognize the problem of data archiving and re-use that FACADE is addressing and are willing to provide immediate feedback on what they need from such archives. By focusing on the practitioners' problems first, we hope to develop archiving practices that will also benefit researchers and instructors as these archives become available.



Mackenzie Smith

Acknowledgements

This project is made possible by a grant from the U.S. Institute of Museum and Library Services. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the IMLS or the U.S. Government.

Mackenzie Smith, Associate Director for Technology

MIT Libraries, 77 Massachusetts Ave, Cambridge, MA USA, kenzie@mit.edu

Mackenzie Smith has overseen the MIT Libraries' use of technology and its digital library research program since 2003. Projects she has directed there include DSpace, MIT's collaboration with Hewlett-Packard Labs to develop an open source digital repository for scholarly research material in digital formats, and SIMILE, working with the Computer Science Lab and the W3C on Semantic Web applications for libraries. Prior to her tenure at MIT she was the Digital Library Program Manager for the Harvard University Library, and has also held positions in the library IT departments of Harvard and the University of Chicago. She holds a BA from the University of Washington, and an MA in Library Information Science from the University of Chicago. Her research interests are applied technology for libraries and academia, and digital libraries and archives in particular.

References

¹ <http://secondlife.com/>

² Gehry Partners, LLP (<http://www.foga.com/>) is the architectural firm owned by Frank O. Gehry, the architect who designed the Stata Center at MIT and provided the initial research data for the FACADE Project. Gehry's building designs are among some of the most geometrically complex in the world, and he is well-known for using sophisticated 3D modeling software in his design process, most notably CATIA from Dassault Systèmes (see reference 4 below).

³ <http://autodesk.com>

⁴ <http://www.3ds.com/>

⁵ <http://bentley.com/>

⁶ <http://www.ptc.com/>

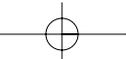
⁷ The Building Information Model (BIM) standard is being developed by the National BIM Standard organization, described here: <http://www.facilityinformationcouncil.org/bim/index.php>.

⁸ A significant prior project on archiving digital design data is reported in *Collecting, Archiving and Exhibiting Digital Design Data*, by Kristine Fallon Associates, Inc., and published in 2004: http://www.artic.edu/aic/collections/dept_architecture/dd.html. The report recognizes the availability and importance of project "outputs" such as client presentations as the data most suitable for long-term archiving, but also recommends keeping other data (including CAD models).

⁹ There are many sources of information on digital preservation tools strategies available. A few organizations that monitor the digital preservation domain are: PADI – Preserving Access to Digital Information – in Australia (<http://www.nla.gov.au/padi/index.html>), the UK Digital Curation Centre (<http://www.dcc.ac.uk/>), the U.S. National Digital Information Infrastructure and Preservation Program (<http://www.digitalpreservation.gov/>), and the Digital Preservation Europe initiative (<http://www.digitalpreservationeurope.eu/>).

¹⁰ For more information on CAD software and data modeling, see Kunwoo Lee's "Principles of CAD/CAM/CAE Systems," Addison-Wesley, 1999.

¹¹ AECnews.com's September 29, 2006 issue described problems incurred by the Airbus company developing its new A380 airplane when they upgraded to CATIA version 5 at different times across the design and manufacturing groups, leading to design incompatibilities and huge financial losses (<http://aecnews.com/articles/2035.aspx>).



Future-Proofing Architectural Computer-Aided Design: MIT's FACADE Project

¹² IGES 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 U.S. Department of Defense as NBSIR 80-1978. It is still widely deployed but is now deprecated in favor of STEP.

¹³ STEP, or 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 profiles” for different types of products or industries.

[http://www.tci84-sc4.org/SC4_Open/SC4%20Legacy%20Products%20\(2001-08\)/STEP_\(10303\)/](http://www.tci84-sc4.org/SC4_Open/SC4%20Legacy%20Products%20(2001-08)/STEP_(10303)/)

¹⁴ 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/>.

¹⁵ 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.

¹⁶ PDF/E is described on the Adobe website <http://www.adobe.com/products/acrobat3d/>.

¹⁷ DSpace is detailed on the project website at <http://dspace.org> and the new architecture is described at <http://wiki.dspace.org/index.php/ArchReviewReport>.

¹⁸ DROID <http://droid.sourceforge.net/wiki/index.php/Introduction>

¹⁹ JHOVE <http://hul.harvard.edu/jhove/>

²⁰ PRONOM <http://www.nationalarchives.gov.uk/pronom/>

²¹ GDFR <http://hul.harvard.edu/gdfr/>

²² METS profiles are described on the project's homepage

<http://www.loc.gov/standards/mets/mets-profiles.html>.

²³ The Resource Description Framework (RDF) is a standard data model from the W3C and is a core part of the Semantic Web infrastructure. RDF can describe any labeled, directed graph so it is useful for complex collection relationship maps such as the Project Information Models created for FACADE <http://www.w3.org/RDF/>.

²⁴ EMC's VMWare product is described on the company's website <http://www.vmware.com/>.

²⁵ QEMU is an open source software machine emulator and virtualizer <http://fabrice.bellard.free.fr/qemu/>.

²⁶ The concept of a “designated community” is a key idea from the Open Archival Information System (OAIS) Reference Model (<http://public.ccsds.org/publications/archive/650xob1.pdf>) which has provided a useful and widely adopted framework and terminology to the digital archives community. The designated community reflects the intended audience of an archive, and helps to specify what knowledge and other expertise the archive can expect of its customers in the future.

LA PÉRENNISATION DES DONNÉES DE CONCEPTION ARCHITECTURALE : LE PROJET FACADE AU MIT

La conception assistée par ordinateur et l'architecture

L'utilisation des logiciels de CAO 3D se généralise dans les agences d'architectes, notamment pour la restauration de bâtiments et la conservation du patrimoine. Ces logiciels sont de plus en plus appréciés dans les écoles d'architecture pour leur usage interactif dans l'enseignement, et permettent également de créer des environnements urbains virtuels tels que Second Life¹. Les maquettes 3D permettent de visualiser de manière dynamique, grâce aux ordinateurs, la géométrie ainsi que les diverses propriétés d'objets physiques. Alors que les architectes plus âgés préfèrent dessiner sur papier et travailler avec des maquettes physiques, les jeunes concepteurs, eux, ne travaillent plus que sur ordinateur, et privilégient la CAO 3D. Quelques grands architectes internationaux tels que Frank Gehry, Moshe Safdie et Zaha Hadid, entre autres, ont adopté depuis quelques années déjà des logiciels de modélisation 3D complexes pour la conception et la construction de leurs projets architecturaux.

À titre d'exemple, l'agence de Frank Gehry², Gehry Partners, LLP, a conçu le Ray and Maria Stata Center, au MIT, à la fin des années 1990 grâce au logiciel CATIA version 4. Ce projet rassemble plus d'une centaine de milliers de fichiers numériques pour une taille dépassant les 140 To. Générées par CATIA, les maquettes 3D originales de ce bâtiment inauguré en 2004 sont déjà obsolètes trois ans plus tard, menaçant de disparition la mémoire institutionnelle de l'évolution de ce projet. Le dossier des ouvrages exécutés (plans 2D) a été archivé mais, vu la complexité géométrique du bâtiment, il ne suffira pas à donner une bonne compréhension du projet dans son intégralité. Un archivage soigneux des différentes données liées à ce projet est essentiel, faute de quoi ces archives deviendront, dix ans à peine après la conception du bâtiment, illisibles et inutilisables.

De nombreux éditeurs proposent aujourd'hui des logiciels CAO aux métiers de l'architecture, de l'ingénierie et de la construction. Parmi les logiciels les plus répandus, on trouve AutoCAD et Revit édités par Autodesk³, CATIA et SolidWorks édités par Dassault Systèmes⁴, Microstation édité par Bentley Systems⁵, et Pro/Engineer édité par Parametric Technology Corporation⁶. Les maquettes 3D sont de plus en plus employées lors des phases de conception, d'ingénierie et de construction d'un projet architectural et sont même parfois utilisées pour la gestion du bâtiment (entretien, rénovation, etc.). Les principes d'utilisation de la CAO 3D pour les projets architecturaux ne sont pas normalisés et peuvent entraîner des erreurs de conversion. C'est pour cette raison que les professionnels de l'architecture, de l'ingénierie

La pérennisation des données de conception architecturale : le projet FACADE au MIT

et de la construction tentent de convenir d'une norme de « Modèle d'information du bâtiment » (BIM)⁷ pour définir le processus d'élaboration d'une maquette 3D standard comprenant toutes les caractéristiques et les informations importantes à communiquer aux diverses parties impliquées dans le projet ainsi qu'aux futurs propriétaires du bâtiment. Certains logiciels récents prétendent offrir une compatibilité avec les BIM, mais la norme n'est pas encore arrêtée, et son contenu exact comme le champ qu'elle recouvrira sont encore mal définis.

Au cours d'un projet d'architecture, il faut archiver de manière durable plusieurs types de fichiers numériques issus du processus de conception et de construction. En plus des maquettes virtuelles 3D, plusieurs centaines voire milliers de fichiers 2D structurés en calques documentent les différentes caractéristiques du bâtiment (la CVC [chauffage, ventilation, climatisation], l'installation électrique, l'aménagement des étages, etc.). En outre, on trouve des dessins ou des croquis du bâtiment présentés aux clients à des étapes clés du projet⁸, ou encore des photos et des vidéos des sites de construction, des sites internet dédiés au bâtiment en cours de construction, et bien d'autres supports multimédia liés au projet. On compte aussi d'abondants échanges entre les architectes, les clients, les prestataires et d'autres parties intéressées sous la forme d'emails, de demandes d'information, d'instructions complémentaires des architectes. Afin de représenter au mieux un projet architectural, il est nécessaire de regrouper et de relier l'ensemble de ces informations numériques entre elles pour en faciliter la recherche, l'exploitation et la pérennisation. Le projet FACADE appelle ces groupements de données numériques les « Modèles d'information du projet » (PIM).

Lors de nos discussions avec des cabinets d'architectes, ils nous expliquent que l'usage de l'informatique dans leur processus de conception les oblige à gérer ces données numériques afin qu'elles soient exploitables et réutilisables pour les besoins internes du cabinet. Comprendre pourquoi telle décision a été prise, quand, et par qui, est aussi difficile aujourd'hui pour ces cabinets que pour les historiens dans les futures décennies.

Il existe aussi de nombreux projets utilisant des maquettes 3D ou des technologies numériques similaires pour aider à la compréhension, à la gestion, à la restauration et à la conservation de bâtiments existants. En Europe tout particulièrement, de nombreux projets de restauration de monuments historiques utilisent une technologie de relevé par balayage laser pour créer des nuages de points et un modèle paramétrique 3D des bâtiments en amont des grands projets de conservation et de rénovation. Cette capacité appréciée par les architectes contemporains de soumettre un édifice à des analyses structurelles ou autres à l'aide d'une maquette 3D devient de plus en plus utile pour les bâtiments anciens. Ces projets patrimoniaux génèrent d'énormes quantités de données numériques, qui doivent être gérées et conservées parallèlement aux données issues des bâtiments en cours de conception.

MacKenzie Smith

Pérennisation des maquettes 3D

La plus grande partie des données numériques issues des projets architecturaux est facile à conserver grâce aux diverses techniques de conservation numérique existantes. Par exemple, les plans DAO 2D générés par le logiciel AutoCAD peuvent être conservés en utilisant un format métier (DWG ou DXF) et convertis pour un archivage à long terme dans un format reconnu comme standard international tel que le PDF/A. Ces plans au format PDF/A peuvent alors être conservés dans l'attente d'une migration vers un format plus pérenne si nécessaire. Des logiciels permettant de telles manipulations existent déjà pour permettre une stratégie standardisée de pérennisation des plans DAO 2D. Pour les autres données issues d'un projet architectural, les images, les vidéos, les communications (emails entre autres), il existe aujourd'hui des techniques de conservation similaires qui peuvent être mises en place⁹.

Par contre, la conservation à long terme de maquettes virtuelles 3D est une tâche bien plus complexe. Les maquettes 3D sont générées par des logiciels propriétaires utilisant des formats de fichiers « natifs » (ou propriétaires). Chaque logiciel utilise sa propre technique pour représenter la forme géométrique de la maquette et de ses propriétés associées, ainsi que les paramètres fournis par l'utilisateur (modèle paramétrique). Ces techniques font appel à des équations mathématiques extrêmement complexes (équations paramétriques B-spline et NURBS, équations non paramétriques, ou une combinaison de ces méthodes)¹⁰. Les logiciels sont donc tous différents les uns des autres, avec des avantages bien distincts pour chacun. Pour ces raisons, les créateurs de solutions logicielles CAO sont peu enclins à définir des formats paramétriques normalisés pour une représentation des données internes, si bien que les maquettes 3D ne peuvent être correctement ouvertes et interprétées que par la version exacte du logiciel qui les a produites. On rencontre même des erreurs considérables de conversion entre les différentes versions d'un même logiciel¹¹. La situation devrait s'améliorer avec l'avènement des Modèles d'information du bâtiment (BIM) normalisés, étant donné les échanges toujours plus nombreux de maquettes virtuelles 3D entre les différents acteurs du secteur de l'architecture et de la construction. Il est cependant difficile de prédire quand les BIM seront devenus le standard d'échange pour toute la profession.

À l'instar des logiciels de jeux vidéo et de réalité virtuelle, il sera à l'avenir nécessaire de pérenniser tous les logiciels originaux et propriétaires utilisés pour créer une maquette virtuelle 3D paramétrique afin de pouvoir intervenir sur ces maquettes. Autrement dit, si les maquettes virtuelles 3D sont archivées dans les formats d'origine, il est nécessaire d'archiver aussi les logiciels propriétaires qui sont à l'origine de ces maquettes. Conserver les logiciels originaux sur le long terme n'est pas facile, et nécessite des techniques complexes d'émulation de systèmes d'exploitation. L'émulation de systèmes matériels et de systèmes d'exploitation

La pérennisation des données de conception architecturale : le projet FACADE au MIT

s'avère très compliquée et très coûteuse, et serait difficilement réalisable dans chacun des centres d'archives conservant des archives d'architecture numérique.

Des normes internationales pour l'exportation des données entre les logiciels de CAO 3D existent, mais elles entraînent toutes des pertes d'information des maquettes paramétriques d'origine puisqu'elles les réduisent à une géométrie et à des caractéristiques statiques (sous forme solide, surfacique ou filaire) et ne prennent pas en compte les paramètres spécifiques définis par l'utilisateur lors de la conception. Malgré ces limites, la majorité des standards de données 3D sont reconnus et arrivent à maturité. Ces standards comprennent les formats IGES¹², STEP¹³, IFC¹⁴, STL¹⁵ et le nouveau format PDF/E¹⁶ d'Adobe. Le projet FACADE s'est concentré initialement sur le standard IFC comme format de référence privilégié pour l'archivage, mais nous nous penchons également sur le format STEP à cause de sa grande diffusion dans de nombreux secteurs ; le format est en passe de devenir obligatoire pour les documents de conception architecturale remis au gouvernement fédéral des États-Unis.

Archivage des données architecturales numériques avec DSpace

Dans le cadre du projet FACADE, les bibliothèques du MIT vont créer un dépôt d'archives prototype pour les données architecturales numériques avec DSpace, une plateforme libre destinée à l'archivage des recherches académiques sous forme numérique¹⁷. DSpace permet la conservation basique (c'est-à-dire au niveau binaire) de tous les formats numériques, ainsi que la conservation fonctionnelle de certains formats courants, par le moyen de techniques de conversion standardisées. Le niveau de compatibilité de la conservation pour un format de fichiers donné est dicté par une procédure interne via un registre de formats. Le responsable de l'archivage doit s'assurer que les formats « compatibles » sont tous soumis à une stratégie de conservation définie et applicable au sein de son dépôt d'archives.

Pour faciliter l'archivage des données numériques d'architecture, le projet FACADE développe la capacité de DSpace à identifier, à valider et à documenter les différents formats (métadonnées techniques pour formats de fichiers). Ce travail se fonde sur des outils existants tels que DROID¹⁸ et JHOVE¹⁹ pour l'identification des formats. DSpace intégrera aussi les registres de formats internationaux (PRONOM²⁰ pour l'instant, et GDFR²¹ dans un futur proche) afin que les stratégies de conservation développées dans DSpace tirent parti des registres de formats tiers pour pouvoir surveiller l'évolution des différents formats existants et identifier les futurs outils de migration.

Le projet FACADE permettra également de créer des gestionnaires de paquets pour les Modèles d'information du projet qui incluront les fichiers CAO/DAO 2D et 3D ainsi que les fichiers multimédias et les données de communication (emails,

Mackenzie Smith

etc.). Un profil METS²² ou une norme de paquetage structurel similaire sera défini dans ce but et des outils d'intégration seront créés pour DSpace afin de permettre l'archivage, la recherche et l'exploitation de ces archives via Internet. Il est peu probable que ces gestionnaires de paquets fonctionnent pour toutes les données issues de projets architecturaux. Il sera nécessaire de renseigner manuellement certaines informations, notamment la nomenclature interne du cabinet d'architecture pour ses fichiers à structure de calques 2D. Nous pensons néanmoins que la création d'une boîte à outils facilitera grandement le travail des archivistes qui reçoivent de grandes collections de données issues de projets architecturaux et qui désirent les intégrer dans leurs archives numériques.

Afin de simplifier la recherche et l'identification des archives numériques ainsi collectées, le projet FACADE créera une description simplifiée (au niveau du PIM) au format Dublin Core. Des métadonnées plus détaillées seront également créées pour certains fichiers importants, notamment pour les maquettes virtuelles 3D, pour certains fichiers de présentation des principaux projets, ou pour des fichiers de communication importants. En plus des métadonnées descriptives, le projet FACADE permettra d'inclure le plus de métadonnées techniques possible lors du processus d'intégration (autant que possible de manière automatisée) et enregistrera ces informations dans DSpace pour un traitement ultérieur. Enfin, des métadonnées structurales sur l'ensemble des fonds seront créées afin de permettre aux utilisateurs de naviguer dans les fonds au moyen d'une interface graphique proposant des liens. Une approche basée sur une ontologie RDF (*Resource Description Framework*) est à l'étude pour que toutes les données du PIM soient reliées entre elles de manière cohérente.

Enfin, le projet FACADE étudie la possibilité de créer un système d'émulation simple pour DSpace dans le but de tester la capacité d'un centre d'archives numériques à faire fonctionner d'anciens logiciels dans un environnement d'exploitation différent. Par exemple, tenter de faire fonctionner CATIA version 4 (le logiciel de CAO/DAO utilisé par le cabinet d'architecte de Frank Gehry lors de la conception du Stata Center du MIT) avec DSpace, sous UNIX. De telles expériences nous apprendront quelles sont les possibilités offertes aujourd'hui par les logiciels de virtualisation et d'émulation prêts à l'emploi (par exemple, VMWare²⁴ ou QEMU²⁵). Cependant, nous pensons qu'il y a peu de chances de trouver des solutions concrètes aux problèmes de l'archivage numérique dans l'immédiat, pour les raisons mentionnées ci-dessus.

Conclusion

Les maquettes 3D prennent une place importante parmi les documents créés par les architectes et les designers d'aujourd'hui, par la recherche et par la rénovation d'édifices anciens, et susceptibles de servir à la recherche historique. Pourtant l'archivage de ces maquettes et de leurs représentations pose aujourd'hui un réel défi. La plupart des centres d'archives numériques évitent de les conserver et

La pérennisation des données de conception architecturale : le projet FACADE au MIT

intègrent seulement les formats de fichiers simples (documents, images, etc.). Or avec ces méthodes d'archivage, une grande partie de l'historique du projet, dépeignant l'intention et la conception originelles de l'architecte, sera perdue. Cet usage des maquettes virtuelles 3D est destiné à s'intensifier avec l'avènement des Modèles d'information du bâtiment, de plus en plus fréquents dans les projets architecturaux. Il faudra donc rendre accessibles toutes les données relatives au projet, de la conception à la construction, sans oublier les fichiers liés au processus de construction en lui-même, si l'on veut avoir une meilleure compréhension des projets architecturaux dans l'avenir.

Le monde de l'enseignement aura également beaucoup à apprendre de l'utilisation des données architecturales numériques. Les étudiants en architecture ne disposent pas à l'heure actuelle de bibliothèque de fichiers de conception numérique. Ils apprennent à maîtriser les outils et les techniques de la conception numérique mais sans avoir accès à des cas d'étude provenant de projets architecturaux concrets. Il n'existe pas d'équivalent de diapo-thèque pour les fichiers de conception numériques. Les étudiants ont un accès limité aux informations décrivant les processus de conception des bâtiments récents, informations contenues dans les Modèles d'information du bâtiment et dans nos Modèles d'information du projet. Dans le cadre du projet FACADE, nous donnerons des exemples d'utilisation des modèles numériques dans le cadre de l'enseignement pour que les écoles d'architecture prennent conscience des avantages liés à l'accès à de telles archives.

Plus concrètement, la « communauté d'utilisateurs »²⁶ (ou le public visé) par le projet FACADE se compose de professionnels de l'architecture et de la conception utilisant quotidiennement des logiciels de CAO/DAO. Les chercheurs et les enseignants pourraient aussi être intéressés par ces archives numériques, mais, vu la rareté de telles archives aujourd'hui, on dispose pour l'instant de peu de retours d'expérience et les besoins sont mal connus. Les architectes, par contre, sont déjà confrontés à la problématique de l'archivage de données et de leur exploitation, et peuvent nous faire savoir rapidement ce qu'ils attendent de telles archives. En nous concentrant tout d'abord sur les problématiques des architectes, nous espérons mettre en place de nouvelles pratiques d'archivage qui, une fois opérationnelles, sauront également répondre aux besoins des chercheurs et des enseignants.

Remerciements

Un grand merci à l'US Institute of Museum and Library Services (IMLS) qui subventionne ce projet. Les propos et les conclusions de ce document n'engagent que leurs auteurs et ne reflètent en aucun cas la position, explicite ou sous-entendue, de l'IMLS ou du gouvernement fédéral des États-Unis.

MacKenzie Smith

MacKenzie Smith, Directrice déléguée à la technologie
MIT Libraries, 77 Massachusetts Ave, Cambridge, MA, États-Unis, kenzie@mit.edu
MacKenzie Smith supervise depuis 2003 l'utilisation de la technologie dans les bibliothèques du MIT ainsi que le programme de recherche sur la bibliothèque numérique. Les divers projets sous sa direction comprennent DSpace, un projet du MIT en collaboration avec les laboratoires Hewlett-Packard dans le but de développer un dépôt d'archives numériques au format libre pour archiver les recherches académiques créées sous format numérique, ainsi que SIMILE, en collaboration avec les laboratoires Computer Science et W3C, pour l'étude des applications du Web sémantique pour les bibliothèques. Avant son arrivée au MIT, Mme Smith était directrice du programme de la bibliothèque numérique à la bibliothèque de l'université d'Harvard. Elle a également occupé des postes à responsabilité dans le département informatique de la bibliothèque de l'université d'Harvard et à celle de l'université de Chicago. Elle a obtenu une licence de l'université de Washington et une maîtrise en bibliothéconomie de l'université de Chicago. Ses recherches se sont portées sur les technologies appliquées aux bibliothèques et aux universités, notamment les bibliothèques numériques et les centres d'archives.

Notes

¹ <http://secondlife.com/>

² Gehry Partners, LLP (<http://www.foga.com/>) est l'agence d'architecture de Frank O. Gehry, architecte à l'origine du MIT Stata Center, et a fourni les premières données de recherches pour le projet FACADE. Les bâtiments conçus par Frank O. Gehry font partie des bâtiments aux formes géométriques les plus complexes au monde. Son agence est reconnue pour l'utilisation de logiciels de modélisation 3D pour la phase de conception de ses projets, notamment CATIA de Dassault Systèmes (cf. note 4 ci-dessous).

³ <http://autodesk.com>

⁴ <http://www.3ds.com/>

⁵ <http://bentley.com/>

⁶ <http://www.ptc.com/>

⁷ La norme des Modèles d'information du bâtiment (BIM) est développée par l'organisation National BIM Standard, décrite plus en détail sur cette page : <http://www.facilityinformationcouncil.org/bim/index.php>.

⁸ Un premier projet portant sur l'archivage des données de conception numériques est étudié dans un ouvrage de Kristine Fallon Associates, Inc., publié en 2004, intitulé *Collecting, Archiving and Exhibiting Digital Design Data* (http://www.artic.edu/aic/collections/dept_architecture/dd.html). Cet exposé estime que les « livrables » issus de projets, tels que les présentations clients, sont les données les plus appropriées pour un archivage pérenne, mais recommande également l'archivage d'autres données (notamment les maquettes virtuelles 3D).

⁹ Plusieurs sources d'information sont disponibles sur les outils nécessaires à la conservation numérique. Voici quelques acteurs dans le domaine de la conservation numérique : Le PADI (Preserving Access to Digital Information) pour l'Australie (<http://www.nla.gov.au/padi/index.html>), le UK Digital Curation Centre (Royaume-Uni) (<http://www.dcc.ac.uk/>), le US National Digital Information Infrastructure and Preservation Program (<http://www.digitalpreservation.gov/>), et l'initiative Digital Preservation Europe (<http://www.digitalpreservationeurope.eu/>).

¹⁰ Pour plus d'information sur les logiciels CAO/DAO et sur les modèles de données, voir *Principles of CAD/CAM/CAE Systems* de Kunwoo Lee, éditions Addison-Wesley (1999).

¹¹ L'article daté du 29 septembre 2006 sur le site AECnews.com évoque les problèmes rencontrés par la société Airbus pendant le développement de son nouvel appareil A380 lors du passage à la version 5 du logiciel CATIA, à plusieurs étapes de la conception et de la construction.

La pérennisation des données de conception architecturale : le projet FACADE au MIT

Ces difficultés ont créé de nombreuses incompatibilités, entraînant d'importantes pertes financières (<http://aecnews.com/articles/2035.aspx>).

¹² Le format IGES a été initialement créé en 1979 par plusieurs vendeurs et utilisateurs de logiciels CAO/DAO avec le soutien du National Bureau of Standards et de l'US Department of Defense dans le cadre de la norme NBSIR 80-1978. Ce format est très répandu mais est en passe de devenir obsolète face au format STEP.

¹³ STEP (Standard for the Exchange of Product) est qualifié norme ISO 10303 et décrit les informations de représentation nécessaires à l'interprétation des données « produites » tout au long de leur cycle de vie. Il est composé de centaines de « parties » décrivant les nombreuses caractéristiques des données, et de « profils d'application » pour plusieurs types de produits ou d'industries.

[http://www.tc184-sc4.org/SC4_Open/SC4%20Legacy%20Products%20\(2001-08\)/STEP_\(10303\)/](http://www.tc184-sc4.org/SC4_Open/SC4%20Legacy%20Products%20(2001-08)/STEP_(10303)/)

¹⁴ Le format IFC (*Industry Foundation Classes*) a été développé par l'International Alliance for Interoperability (IAI) pour l'industrie de l'architecture, de l'ingénierie et de la construction. Pour plus d'information, voir <http://www.ifcwiki.org/>.

¹⁵ Le format STL (la stéréolithographie) a été développé dans les années 1980 par 3D Systems, une société qui fournit des machines de prototypage rapide pour des objets industriels (ou produits). Ce format est largement utilisé dans l'industrie de la CAO/DAO 3D notamment pour les géométries de surface.

¹⁶ Le format PDF/E est décrit plus en détail sur le site d'Adobe <http://www.adobe.com/products/acrobat3d/>.

¹⁷ Pour plus de détail sur DSpace, voir le site internet du projet <http://dspace.org>, et, pour la nouvelle architecture, <http://wiki.dspace.org/index.php/ArchReviewReport>.

¹⁸ DROID <http://droid.sourceforge.net/wiki/index.php/Introduction>

¹⁹ JHOVE <http://hul.harvard.edu/jhove/>

²⁰ PRONOM <http://www.nationalarchives.gov.uk/pronom/>

²¹ GDFR <http://hul.harvard.edu/gdfr/>

²² Les profils METS sont détaillés sur la page d'accueil <http://www.loc.gov/standards/mets/mets-profiles.html>

²³ Le format RDF (*Resource Description Framework*) est un modèle de données standardisé piloté par le consortium W3C et est au cœur de l'infrastructure du Web sémantique. Le format RDF peut décrire n'importe quel modèle de graphique et est donc très utile pour décrypter les modèles complexes tels que les Modèles d'information du projet créés dans le cadre du projet FACADE <http://www.w3.org/RDF/>.

²⁴ La solution VMWare développée par EMC est décrite plus en détail sur le site suivant : <http://www.vmware.com/>.

²⁵ QEMU est un logiciel libre qui permet de nombreuses possibilités d'émulation et de virtualisation : <http://fabrice.bellard.free.fr/qemu/>.

²⁶ Le concept de « communauté d'utilisateurs » est l'idée centrale du modèle de référence OASIS (*Open Archival Information System*) (<http://public.ccsds.org/publications/archive/650xob1.pdf>) qui s'est avéré d'une grande aide en apportant un cadre et une terminologie unifiée dans le domaine de l'archivage numérique. La « communauté d'utilisateurs » représente le public visé par un service d'archives, et la définition de ce public permet de cerner les attentes en termes de connaissances et d'expertise des clients du service d'archives.