Preservation and Access of Born-Digital Architectural Design Records in an OAIS-Type Archive

Timothy Walsh
LIS 500: Independent Study
Simmons College School of Library and Information Science
Spring 2015

Advisor: Dr. Katherine M. Wisser

Table of Contents

1 | Introduction

"Architecture is ... one of the languages used by society to express its aspirations, reflect its psyche, and respond to its social and economic needs. The records of architecture and of the built environment should therefore be collected with a conscious regard to its role as a manifestation of other cultural forces."[1]

Although libraries, archives, and museums in North America and Europe have collected, preserved, and provided access to architectural design records for more than a century, significant changes in architectural modes of production following the introduction of Computer Aided Design (CAD) technology in the late twentieth century have significantly complicated this mission. This paper, the result of a self-directed independent study undertaken at the Simmons College School of Library and Information Science during the Spring 2015 semester, addresses the issue of how cultural institutions might provide long-term preservation and access of born-digital architectural design records such as CAD models. The first part of this paper provides some background on the development of these technologies, their complicating features, and archival literature and projects addressing this topic to date. In its second part, the paper looks at how these files might be preserved and made accessible in a digital archive through examination and application of the Open Archival Information System (OAIS) reference model, an international standard for digital stewardship.

1 Nicholas Olsberg, “Documenting Twentieth-Century Architecture: Crisis and Opportunity,” *The American Archivist* 59, no. 2 (1996): 130.

2 | Background

2.1 Brief History of CAD/BIM

The first CAD program, Sketchpad, was designed by Ivan E. Sutherland in 1961-1962.[2] Developed as part of Sutherland's PhD thesis at MIT, Sketchpad utilized a now novel-seeming input mechanism: “the designer interacted with the computer graphically by using a light pen to draw on the computer's monitor.”[3] Shortly thereafter, aerospace and automotive companies such as General Motors, Ford, Lockheed, and Marcel Dassault created their own 2D drafting applications.[4] In the mid-1960s, Control Data Corporation released the first commercially available CAD system, and a number of rivals soon followed suit. These early systems were defined by prohibitively high prices and a reliance on large and expensive mainframe computers.[5] The 1970s continued to see heavy development in 2D CAD programs, as well as the development of a few standards, including IGES (Initial Graphic Exchange Standard), a “widely-used data-transfer format in CAD software” still in use today.[6]

The advent of personal computers in the 1980s introduced a new potential market for CAD software vendors. Autodesk, the first vendor to capitalize on this market, released AutoCAD in 1982.[7] Although initially less powerful than its competitors, because AutoCAD ran as a desktop application on PCs and was designed from the outset to be sold at a much lower

---

2 Kurt Helfrich, ”Questions of authenticity: challenges in archiving born-digital design records,” *Art Libraries Journal* 35, no. 3 (2010): 23.
3 CADAZZ, “CAD software history, 1960s,” *CAD Software – history of CAD CAM*, 2004, https://web.archive.org/web/20150207080928/http://cadazz.com/cad-software-history.htm
4 Wayne Carlson, “Section 10: CAD/CAM/CADD/CAE,” in *A Critical History of Computer Graphics and Animation* (2003).
5 CADAZZ, “CAD software history, 1960s.”
6 CADAZZ, “CAD software history, 1970s.”
7 Carlson, “Section 10: CAD/CAM/CADD/CAE.”

price, it enabled small and medium-sized design firms to take advantage of computer drafting for the first time.[8] AutoCAD's native file formats, DWG and DXF, have since become the "default standard for CAD packages,"[9] and by 1990, Autodesk had sold half a million copies of the product.[10]

The 1980s and 90s also saw two significant changes in CAD systems: the advent of 3D modeling and the development of 32-bit operating systems for PCs. Dassault's CATIA and Parametric Technology Corp (PTC)'s Pro/Engineer (which reintroduced use of the light pen) were among the first commercially successful 3D CAD products in a market that had exceeded revenues of $1 billion as early as 1981.[11] By 1994, Autodesk had followed suit by introducing 3D capabilities in AutoCAD, taking advantage of Windows NT, the new 32-bit Windows operating system for PCs.[12] Other competitors, including SolidWorks, Bentley Systems, and CADKEY, likewise took advantage of the developments in PC processing power, graphics performance, and functionality, solidifying PCs as the tool of choice for working architects.

With the development of robust 3D modeling, CAD models and systems firmly entrenched their importance to the practice of architecture. Alex Ball writes,

> The move to three dimensions was the point at which CAD models stopped being mere conveniences for drawing blueprints and started taking on importance in their own right. With 3D models, it became possible to design shapes that could not be clearly or adequately expressed by three 2D elevations. The ability to analyze designs in 3D meant that more ambitious designs could be attempted, and also that standards for design checks were raised beyond what could be done by eye. In the context of industrial product design, 2D surrogates soon became inadequate records and

---

8 Helfrich, "Questions of authenticity," 23.
9 Carlson, "Section 10: CAD/CAM/CADD/CAE."
10 Marian Bozdoc, "The History of CAD: 1990-1991," iMB, 2003, http://mbinfo.mbdesign.net/CAD-History.htm
11 CADAZZ, "CAD software history, 1986-1989."
12 CADAZZ, "CAD software history, 1990-1994."

regarded as dangerously open to misinterpretation.[13]

The affordances of 3D CAD systems changed architectural practice and education in profound ways. In his prescient talk at the 1999 Massachusetts Committee for the Preservation of Architectural Records (Mass COPAR) program “Blueprints to Bytes: Architectural Records in the Electronic Age,” William Mitchell outlined several of the novel features of 3D CAD systems that been made possible by advances in software sophistication and available computing power since the 1990s. Namely, 3D CAD systems:

- Enable the user to generate perspective renderings and animations as reports from the underlying database;
- Produce construction documents such as sections and elevations as well as traditional 2D architectural drawings on demand;
- Run analyses such as thermal simulations, structural simulations, and fluid dynamic models to simulate the effects of airflow on a structure;
- Produce 3D physical models using computer-aided manufacturing and widely available 3D printers; and
- Integrate with construction processes via CAD/CAM processes to drive fabrication machinery.[14]

3D CAD systems offer an unprecedented number of possibilities for use and reuse of data, much like a musical score from which “you can produce many different performances by the

13 Alex Ball, *Preserving Computer-Aided Design (CAD)*, DPC Technology Watch Report 13-02 (Digital Preservation Coalition, 2013): 5.

14 William Mitchell, “New Digital Technologies in Architecture and their Implications for Architectural Records,” in *Blueprints to Bytes: Architectural Records in the Electronic Age* (Boston: Massachusetts Committee for the Preservation of Architectural Records, 1999): 3-4.

application of appropriate software and machinery."[15]

Recent years have seen yet another important development in computer aided design: Building Information Modeling (BIM). BIM products "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"[16] by representing "both the physical and intrinsic properties of a building as an object-oriented model tied to a database."[17] In practice, this means that BIM software such as ArchiCAD, CATIA, Digital Project, and Revit natively encourage standardization of parts and allow "linking of images with technical and construction data documenting a building's entire life cycle, from construction to operation, that is ideal for the benefits offered by the internet and remote design."[18] The International Foundation Class (IFC) file format and viewing software such as Navisworks, which "allows for data collection, construction simulation and clash detection," have been developed to promote interoperability and limit data loss that arises when exchanging files between proprietary software systems. As a result, The Business Value of BIM in North America study found that 70% of architects and 74% of contractors were using BIM systems as of 2012.[19]

The impact that the new functionalities of 3D CAD and BIM systems have had on architectural practice is hard to overstate. Bernhard Franken's Bubble,[20] Greg Lynn's

---

15 Ibid., 7.

16 MacKenzie Smith, "Curating Architectural 3D CAD Models," *The International Journal of Digital Curation* 1, no. 2 (2009): 99.

17 Michael Bergin, "A Brief History of BIM," *archdaily*, December 7, 2012, http://www.archdaily.com/302490/a-brief-history-of-bim/

18 Helfrich, "Questions of authenticity," 23-24.

19 Bergin, "A Brief History of BIM."

20 Bernhard Franken, Berthold Scharrer, and Inge Wolf, "Save the Bubble," in *Architecture and Digital Archives* (Gollion: Infolio, 2008): 161.

Embryological House[21], Frank Gehry's Bilbao Guggenheim[22], and numerous other examples of so-called “blob architecture” would have been extremely difficult or even impossible had these architects been forced to rely on 2D drafting alone to design and construct their projects. More broadly, 3D CAD software has become "the design tool of choice" for the architectural field, used in architectural programs, small firms, and large firms alike,[23] effectively replacing in one generation a set of two-dimensional, paper-oriented tools and techniques that had been commonplace for several centuries prior.

2.2 Complicating Features and Properties of CAD/BIM Systems

A number of features of the CAD/BIM market, its competing software systems, and their use by architects in practice introduce complexities into the task of preserving and providing access to these cultural products. To date, the most thorough exploration of this issue is Alex Ball's *Preserving Computer-Aided Design (CAD)*, a Digital Preservation Coalition Technology Watch Report released in 2013. The following section, which pulls heavily from Ball's work, summarizes some of these complicating features.

2.2.1 Market forces

As the above history suggests, the computer aided design market has from its outset been fast changing and highly competitive, “resulting in CAD systems that are ephemeral and largely incompatible with each other.”[24] Because software vendors do not wish to lose market

---

21 Howard Shubert, “Preserving Dgiital Archives at the Canadian Centre for Architecture: Greg Lynn's Embryological House,” in *Architecture and Digital Archives* (Gollion: Infolio, 2008): 255.
22 Mitchell, “New Digital Technologies in Architecture,” 5.
23 Smith, “Curating Architectural 3D CAD Models,” 99.
24 Ball, *Preserving Computer-Aided Design (CAD)*, 2.

share to their competitors, there is little economic incentive for vendors to develop robust export functionalities or reveal details about their proprietary file formats. Even if vendors were to share these details, they may quickly become irrelevant due to rapid software development:

> In order to maintain a competitive edge, there is constant pressure on CAD vendors to release new versions of their software with increased functionality or fewer limitations. Not only does this create instability regarding file formats and their interpretation, it also means that individual versions of CAD packages can become obsolete rather quickly, especially when compared to the required lifespan of the CAD models they create. To put this in concrete terms, a new version of a typical CAD system might be released every six months, and the system withdrawn entirely after ten years.[25]

This rapid software development is exacerbated by the fact that CAD software licenses are often time-limited and new versions of software are not always designed to be backwards-compatible, forcing architectural firms to upgrade their software regularly and frustrating efforts by archivists to access files created before the last few product releases.[26]

2.2.2 Complexity

Interoperability between 3D CAD systems and even between different versions of the same system is further limited by the complexity of the interactions between CAD files and the highly complex, diverse, and proprietary modeling kernels of software systems. In order to write and render complex 3D data, systems use a wide range of "very complex mathematical techniques, for example parametric B-Spline or NURBS equations, non-parametric equations, or a combination of both."[27] Because these techniques play such a significant role in the

25 Ibid., 10
26 Ibid.
27 Smith, "Curating Architectural 3D CAD Models," 101.

rendering of a CAD model, “CAD files tend not to be exhaustive descriptions of a model, but rather more like recipes for building the model within the software.”[28] Because even small changes in geometric representation can have large effects on modeling, “computer aided design (CAD) data translation, especially solid model translation, has been a challenging problem for both industry and academia.”[29] This is a lesson the commercial world has learned, when, for example, delays in Airbus A380 airliner production in the mid-2000s caused by facilities using different versions of CATIA cost Airbus an estimated 4.8 billion Euros.[30] [31]

2.2.3 Layers

Like Photoshop and many other graphically-centered software systems, CAD and BIM systems utilize layers to structure and selectively display data. The use of layers is so commonplace that “conventions on what information to include in each one and how to name them are the subject of national and international standards.”[32] Thus, in order to fully capture and represent an architect's working practice, archivists working with CAD files must determine what conventions were used in the file's creation as well as whether and/or how to allow future users to hide, display, and combine layers as would have been possible at the moment of creation.

---

28 Ibid.
29 Jianchang Qi and Vadim Shapiro, “Geometric Interoperability with Epsilon Solidity,” *Journal of Computing and Information Science in Engineering* 6 (2006): 213.
30 Ball, *Preserving Computer-Aided Design (CAD)*, 10.
31 Nichola Clark, “The Airbus saga: Crossed wires and a multibillion-euro delay,” New York Times, December 11, 2006.
32 Ball, *Preserving Computer-Aided Design (CAD)*, 8.

2.2.4 External references

It is not unusual for architects to import external reference files (called Xref files in AutoCAD) into their CAD models as part of the design process, for instance “a common drawing detail or a drawing template with notes that would be used as a block and be common to multiple drawings.”[33] These files are often stored externally to the CAD file that references them, creating additional dependencies that must be captured and stored in an archival environment. CAD files may also link to each other, in the case of complex projects that are broken into several discrete files, or to databases through an application programming interface (API).[34] Archivists will have to decide whether to retain these external references. In cases where it is deemed necessary, doing so may require activities such as changing links from absolute to relative paths and/or preserving software that enables the interaction between files.[35]

2.2.5 Working practices of architects and architectural firms

It would appear that few architectural firms follow consistent practice when it comes to records management, file naming, and other organizational behaviors that ease the tasks of identification, arrangement, and description for archivists. As Gerald Beasley and Annemarie van Roessel of the Avery Architectural & Fine Arts Library at Columbia University explain, small and medium-sized architectural firms

> serve clients with pressing needs, neither side having time or incentive to be concerned about a long-term legacy beyond the legal records retention regulations … Generally speaking, it is rare for this group to employ a professional records manager,

33 Elys John, “Understanding layers,” in *CAD fundamentals for architecture* (London: Laurence King Pub, 2013).
34 Ball, *Preserving Computer-Aided Design (CAD)*, 12.
35 Ibid.

> and far rarer to employ an archivist. This group, of course, forms the bulk of architectural practices and will generally continue to pass by institutional collectors.[36]

Likewise, larger firms who do have records managers or archivists in their employ typically maintain archives to satisfy legal requirements, giving little thought to long-term preservation because "for many of these architects the very notion of spending money on the past is hard to rationalize, when the future – the next commission, the next competition – is where resources must be invested first."[37]

This general lack of organization often extends to digital objects created during the design process, as the Art Institute of Chicago learned:

> The collaborative nature of Garofalo's practice means that several people may be working on the same project at once. In the absence of strict naming or organizational conventions, files get created with inconsistent or conflicting file names that are often no help in determining the content of the file. In addition, the "final" version of any given design was often not called out as such; several very closely related iterations may be saved in the same directory with no indication of which one was used for the final rendering.[38]

When placed within a context where architects may be porting data back and forth between multiple CAD systems and only retaining software licenses for as long as versions of software are useful,[39] one can begin to understand how architects' organizational practice complicate efforts to describe, preserve, and provide access to their digital design files.

---

36 Gerald Beasley and Annemarie van Roessel, "Digital Architectural Archives in North America," in *Architecture and Digital Archives* (Gollion: Infolio, 2008): 281.

37 Ibid.

38 Kristine Fallon and Carissa Kowalski Dougherty, "A Pilot Project for Born-Digital Architecture Data at the Art Institute of Chicago," in *Architecture and Digital Archives* (Gollion: Infolio, 2008): 386.

39 Ibid.

2.3 Archival literature and projects to date

Professional awareness of the difficulties of preserving born-digital architectural records is evident in archival literature dating back to the 1990s. Several articles in the 1996 special issue of *The American Archivist* on architectural records make mention of the challenges presented to archivists by records in CAD formats. Nicholas Olsberg cited volume and irretrievability of data, unstable storage media, lack of standards, and lack of expertise on the part of archivists as factors that will limit cultural institutions' ability to preserve and provide access to electronic records generated by architects and their firms.[40] In the same issue, William Mitchell, three years before his Mass COPAR speech, likewise explained the volatility of storage media, the tendency to "retain everything in a disorganized way" in digital environments, and software and hardware dependencies of CAD systems as significant obstacles to preservation efforts.[41]

Nonetheless, a number of projects have sought to address these issues and create digital architectural archives in the intervening years. A number of these projects were presented at the Architecture and Born-Digital Archives conference held in Paris, France, in 2007 as part of the Gau:di (Governance, Architecture and Urbanism: a Democratic Interaction) program, the proceedings of which were subsequently published as the book *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*.[42] The projects described in *Architecture and Digital Archives*, including the Art Institute of Chicago's Digital Archive for Architecture System (DAArch), Norway's National Museum of Art,

---

40 Olsberg, "Documenting Twentieth-Century Architecture," 133.
41 William Mitchell, "Architectural Archives in the Digital Era," *The American Archivist* 59, no. 2 (1996): 203.
42 *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*, eds. David Peyceré and Florence Wierre (Gollion: Infolio, 2008).

Architecture, and Design's “E-archive Snøhetta AS”, Deutsches Architekturmuseum (DAM)'s “BMW Bubble” preservation project, and the Canadian Centre for Architecture (CCA)'s *DOCAM* pilot project to preserve Greg Lynn's Embryological House, cited many of the same difficulties presaged by Olsberg and Mitchell a decade before.

Perhaps the highest profile projects related to preserving born-digital architectural records in the past decade have been MIT's FACADE (Future-proofing Architectural Computer-Aided Design) and MIT and Harvard's collaborative FACADE2. The first FACADE project, conducted between 2006 and 2009, included work on format identification for 3D CAD file formats, recommendations for preservation file formats, and the development of a Project Information Model (PIM) ontology for organizing records relating to a given design project.[43] FACADE2 focused on development of a curator's workbench for implementing FACADE's PIM ontology, but “progressed more slowly than expected because, once reviewed by the technical support team, unexpected functional and technical bugs were discovered during testing of the latest version.”[44] Notably, the introduction to the project's final report asserts that, “Archivists at many organizations are dealing with the reality of CAD (Computer Aided Design) files that cannot be made fully accessible because strategies for long-term access of these software-dependent digital objects are not yet mature.”[45]

This assertion appears to be consistent with the findings of a study conducted by Anne R. Barrett as part of her Master's paper for the M.S. in LS degree at the University of North Carolina in 2012. Interviews with twenty-five archival professionals from institutions that

43 Smith, “Curating Architectural 3D CAD Models.”
44 “Final Report: FACADE2: MIT and Harvard Collaboration,” Harvard Library Lab, https://osc.hul.harvard.edu/liblab/sites/default/files/325_final_2013_0.pdf
45 Ibid.

collect architectural records revealed that none of the institutions were actively collecting born-digital architectural records, and a number of these institutions would not accept the records even if offered.[46] There was, however, some hope of change on this front:

> The Art Institute of Chicago, Harvard University's Frances Loeb Library, the Smithsonian, the Northwest Architecture Archives at the University of Minnesota, the University of Florida Smathers Libraries, the Institut francais d'architecture, and the Canadian Center for Architecture are all in the process of building actionable workflows for born-digital architectural records. The processes are continually being re-assessed and re-adjusted as more materials are transferred over or accessioned by these institutions.[47]

At the time of the study, it is worth noting that some of these institutions, such as Harvard University's Frances Loeb Library, only accepted files that had already been migrated to standard document or image formats such as PDF, JPG, and TIF.

As of 2013, the situation seemed much the same. As Jakob Beetz, Stefan Dietze, René Berndt, and Martin Tamke reported in their report on the “DURAARK – Durable Architectural Knowledge” project,

> Currently, no existing approach is able to provide a secure and efficient long-term preservation solution covering the broad spectrum of 3D architectural data, while at the same time taking into account the demands of institutional collectors like architecture libraries and archives as well as those of the private sector including building industry SMEs, owners, operators, and public stakeholders.[48]

---

46 Anne Barrett, “Born-digital Architectural Records: Defining the Archivable Record” (master's thesis, University of North Carolina at Chapel Hill, 2012), 18.

47 Ibid., 20.

48 Jakob Beetz, et al., “Towards the Long-Term Preservation of Building Information Models,” *Proceedings of the 30th CIB W78 International* Conference, Beijing, China, October 9-12, 2013, 209.

3 | An OAIS-Type Digital Architectural Archive

The Open Archival Information System, or OAIS, is an international standard (ISO 14721) initiated by the Consultative Committee for Space Data Systems (CCSDS) and used by libraries, archives, museums, commercial entities, governments, and other agents as a high-level reference model that "identifies and describes the core set of mechanisms by which an OAIS-type archive meets its primary mission of preserving information over the long term and making it available to the Designated Community."[49] The OAIS reference model introduced and defined a number of concepts, including the "information package" and "Designated Community." The reference model also enumerates a list of mandatory responsibilities for OAIS archives:

- "Negotiate for and accept appropriate information from information producers;
- Obtain sufficient control of the information in order to meet long-term preservation objectives;
- Determine the scope of the archive's user community;
- Ensure that the preserved information is independently understandable to the user community, in the sense that the information can be understood by users without the assistance of the information producer;
- Follow documented policies and procedures to ensure the information is preserved against all reasonable contingencies, and that there are no *ad hoc* deletions;

---

49 Brian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide* (2nd Edition), DPC Technology Watch Report (Digital Preservation Coalition, 2014): 2.

- Make the preserved information available to the user community, and enable dissemination of authenticated copies of the preserved information in its original form, or in a form traceable to the original."[50]

Perhaps the most widely-referenced contribution of the reference model is the OAIS functional model, "which defines six high-level services, or functional entities, that collectively define the OAIS's preservation and access operations: Ingest, Archival Storage, Data Management, Preservation Planning, Access, and Administration."[51] In the sections that follow, discussion of how 3D CAD files and other complex born-digital architectural records can be accessioned, preserved, arranged, described, and made accessible are loosely mapped to these six OAIS functional entities. It is the author's hope that by doing so, a picture for management of Computer Aided Design records, and thus an outline of a practical and comprehensive appraisal-to-access workflow, might emerge.

3.1 Ingest

3.1.1 OAIS definition

The Ingest Functional Entity in an OAIS-type archive represents "the set of processes responsible for accepting information submitted by Producers and preparing it for inclusion in the archival store."[52] These processes include Receive Submission, Quality Assurance, Generate AIP, Generate Descriptive Information, and Coordinate Updates.[53] In less technical terms, this step involves accessioning data, validating that the data is authentic and complete,

---

50 Ibid., 7.
51 Ibid., 2.
52 Ibid., 12.
53 *Reference Model for an Open Archival Information System (OAIS) Magenta Book* (Consultative Committee for Space Data Systems, 2012): 4-6 - 4-7.

transforming the data into a preservation-friendly form, and creating descriptive metadata for use in discovery systems.

3.1.2 Importance of active and early appraisal/transfer

As with other born-digital materials, the benefits of early and active appraisal and transfer of records from architects and architectural firms cannot be overstated. Selection and transfer of data is most likely to be successful when archivists can consult with records creators. If files are appraised and transferred within a few years of creation, records creators are more likely to remember and share crucial information about their creation, modification, use, and organization, allowing archivists to record more thorough and accurate contextual information for users of the archive. In cases were file naming conventions and CAD style manuals were absent or inconsistently applied, the creators may be able to identify working and final versions of designs and other important qualities of files. Furthermore, early appraisal and transfer leaves open the possibility of managed migration of CAD files to output or standard file formats by taking advantage of the relatively narrow window of time in which a firm will still have a legally licensed installation of the version of the software used to create them, limiting the need for difficult and time-intensive digital archaeological recovery.

3.1.3 *Data transfer*

Regardless of when data is transferred, it is crucial that archivists do not alter files or their metadata during the transfer process. Here, an understanding of digital forensics and the materiality of digital information proves useful. At the very least, archivists responsible for receiving digital data must be aware that Locard's exchange principle, which states that every contact leaves behind a trace, is “more, not less, true in the supposedly virtual confines of computer systems."[54] As Matthew Kirschenbaum, Richard Ovenden, and Gabriela Redwine explain in the 2010 CLIR report *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*,

> ... [b]orn-digital materials can be compromised not only physically (e.g., broken or exposed to adverse conditions), but also at the logical level (e.g., altered files and metadata). The time between when born-digital materials leave a creator's possession and when they arrive at the repository is marked by a particular vulnerability.[55]

In order to ensure the authenticity of data, archivists can use write blockers – software or hardware devices which prevent accidental over-writing of data on storage media.[56] Archivists must also get in the habit of generating and routinely verifying cryptographic hashes (checksums) for all digital files to ensure that their bitstreams remain consistent over time. Whether files are transferred to an archive on removable media or via network transfer, comparing checksums pre- and post-accession allows the archive to demonstrate the authenticity of their digital holdings.[57] Due to the size and complexity of 3D CAD files and other born-digital architectural records, it is essential that repositories be able to demonstrate their fixity to Producers and Consumers alike in a reliable manner.

Disk imaging is one method of creating bit-level captures of digital data. A disk image is a one-for-one copy of digital data as it resides on a particular storage medium. Long used in computing in general, disk imaging is increasingly becoming a central part of collecting institutions' digital archives workflows:

> Disk images can play an essential role in the acquisition and management of digital collections. Preserved disk images can be used at a later time to provide proof of file

---

54 Matthew Kirschenbaum, Richard Ovenden, and Gabriela Redwine, *Digital Forensics and Born-Digital Content in Cultural Heritage Collections* (Council on Library and Information Resources, 2010): 6.
55 Ibid., 28.
56 Ibid., 21.
57 Ibid., 38.

> integrity and chain of custody. Disk images can ensure continued access to information in collections without depending on physical carriers, which may be fragile or become obsolete; can serve as baselines for comparison when evaluating digital preservation actions; and can provide fail-safe mechanisms (backups) when curatorial actions make unexpected changes to data. Disk images can be shared with other institutions. Finally, disk images provide access to potentially valuable data that resides below the user-accessible portions of the file system, including metadata, recoverable sectors and configuration information.[58]

Research, largely centered around the Bitcurator project, is currently being undertaken to develop workflows that utilize disk images along every step of the digital preservation process, from ingest to storage to access.[59] Although it remains to be seen if this type of complete workflow would be effective for an architectural archive providing access to 3D CAD data, disk images are at the very least an effective tool for accessioning and storing copies of original, unmodified data.

3.1.4 Digital archeology

Regardless of efforts to practice active and early appraisal, archives can expect to receive digital data comprised of obsolete, long-neglected files on obsolete, long-neglected storage media. There are a number of reasons to expect such long-delayed transfers, including the archival profession's reluctance to to accept and preserve electronic design files and the tendency of architects to not to think of long-term preservation as "a major priority for the firm," subordinating these efforts to the demands of day-to-day production.[60]

---

58 Kam Woods, Christopher Lee, and Simson Garfinkel, "Extending Digital Repository Architectures," *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital* Libraries, Ottawa, Ontario, Canada, June 13-17, 2011, 59.

59 Sunitha Misra, Christopher Lee, and Kam Woods, "File-level Access to Disk Images within a Web Browser," *code4lib Journal* 25 (2014).

60 Kathryn Pierce, "Collaborative Efforts to Preserve Born-Digital Architectural Records: A Case Study Documenting Present-Day Practice," *Art Documentation: Journal of the Art Libraries Society of North America* 30, no. 2 (2011): 47.

When confronted with data on obsolete storage media, the first step is to extract the data from the media to a more stable location, such as a backed-up network drive. The ease of this step depends significantly on institutional resources and the age and popularity of the media in question. While 3.5” and 5.25” floppy drives, Zip drives, and other hardware required to read commonly-found obsolete media can still be purchased second-hand, reliable drives of these types have not been manufactured for years. Additionally, lack of compatibility with modern motherboards may require the use of floppy controllers such as the FC5025 (for 5.25” floppy disks) and Kryoflux (for 3.5” and 5.25” floppy disks). Older hard drives, such as those using the Modified Frequency Modulation (MFM) encoding scheme, may likewise necessitate the acquisition of controller cards. As with any transfer, archivists should take to avoid altering data by using write blockers and creating disk images.

Once data has been safely transferred from its original media, the task of identifying and analyzing file contents remains. A number of tools, including JSTOR/Harvard Object Validation Environment 2 (JHOVE2) and Digital Record Object Identification (DROID), are available to archivists who wish to identify and characterize unknown files.[61] Because these tools rely upon file format registries such as the UK National Archives' PRONOM, their effectiveness for CAD formats will depend on the degree to which the formats are documented in the registries. When identification tools produce no or limited results, hex editors can be used to examine file headers for relevant information, and in hard drives used in PCs running Windows operating systems, the registry can be a trove of useful information, including “user and password data, as well as information on which programs have been

61 Julianna Barrera-Gomez and Ricky Erway, *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house* (Dublin, Ohio: OCLC Research, 2013): 24.

installed and uninstalled, and what documents have been opened and closed."[62]

3.1.5. Format normalization

Whether CAD files accepted by an archive are current or obsolete, the archive will need to decide how to preserve them for the long term. It is advised that archives always retain a copy of the original data in its original format, as a means of guaranteeing authenticity and to leave options open for future emulation and migration projects. Archives will have to decide whether to also store migrated 'preservation' versions of files alongside their original bitstreams, and if so, in which formats.

Normalization of 2D CAD files should not prove overly difficult for architectural archives, so long as a software package capable of reading the files can be secured. Industry standard formats such as AutoCAD's DWG, despite being proprietary, can be rendered using a range of available software. As an openly documented format, AutoDesk's DXF offers another option to archivists, although there is "reported to be insufficient detail in some areas of the specifications from version R13 onwards to allow full implementation."[63] If interactive layer functionality is not required, DWG and DXF documents can be converted into open document and image preservation formats such as PDF/A and uncompressed TIFF. These features led those involved with the FACADE project to conclude that "standard formats and preservation strategies can reasonably be applied" to 2D CAD drawings.[64]

For 3D CAD files, Alex Ball recommends that archives "normalize CAD models to at

62 Ciaran Trace, "Beyond the Magic to the Mechanism: Computers, Materiality, and What it Means for Records to Be 'Born Digital'," *Archivaria* 72 (2011): 23.
63 Ball, "Preserving Computer-Aided Design (CAD)," 18.
64 Smith, "Curating Architectural 3D CAD Models," 100.

least one, but ideally two or three, vendor-neutral standard formats," particularly those defined by the STEP (Standard for the Exchange of Product Model Data) international standard (ISO 10303).[65] The FACADE project recommends that four versions of CAD files be kept for preservation: the original; a display format, such as 3D PDF; a "heavyweight" standard format, such as IFC or STEP; and a "lightweight" dessicated format, such as IGES, which retains the simple geometry of the model:

> The rationale it gives for this is as follows. The original format provides the fullest amount of information about the design, but is only usable so long as the original software is actually and legally available (in the short term in most cases). The purpose of the standard formats is to preserve the design information in a vendor-neutral way, portable way. A heavyweight standard is used to preserve the most information possible, accepting the risk that some information may be poorly converted, leading to an inauthentic expression of the design. A lightweight standard is used to preserve a restricted subset of the information (specifically shape data), in the expectation that this subset will be translated robustly and could therefore be used as a fall-back option should the information encoded using the heavyweight standard prove unreliable. The visualization format was chosen to allow convenient display of the model in-browser, using software that is near-ubiquitous among users of the archive.[66]

At present, the two most promising “heavyweight” standards for vendor-neutral 3D CAD file formats are STEP and IFC. Due to their promise as preservation standards, each is worth exploring in some depth.

STEP was first conceived in 1984 as a more fully-featured and rigorously defined replacement for existing standards such as IGES, SET, and VDA-FS.[67] Development has continued since, with elements of the standard being published throughout the 1990s and 2000s. Ball writes,

> The most widely known and widely implemented parts of STEP are AP 203,

65 Ball, “Preserving Computer-Aided Design (CAD),” 2.
66 Ibid., 22.
67 Ibid., 14.

> 'Configuration Controlled 3D Designs of Mechnical Parts and Assemblies', and Part 21, 'Clear Text Encoding of the Exchange Structure', which together define a CAD file format suitable for exchange and archiving known as an AP 203 STEP file (or STEP physical file).[68]

An additional part, AP 242, or 'Managed Model-Based 3D Engineering', is also in development, with additional functionality for “shape data quality information, semantic 3D product and manufacturing information (PMI), approximate geometry (for visualization) and access rights management.”[69] Efforts are likewise in development through the CAx Implementer Forum to test CAD conversion tools for STEP compliance, which should help to increase adoption and limit the amount of information loss that is known to occur as a result of migration between 3D CAD file formats.[70]

The IFC (Industry Foundation Classes) data has been collaboratively developed as a vendor-neutral standard format for BIM product data. IFC has seen widespread adoption as well as development of complementary standards such as National BIM Standard – United States (2012), “which specifies how BIMs, and by extension how standards like IFC, should be used.”[71] These qualities have led some to describe IFC as a “very attractive and truly archival format.”[72]

Regardless of which format is chosen, it must be recognized that migration between 3D CAD formats will necessarily entail some degree of information loss. The most likely culprit for loss is parametric data, as despite work to include support for parametric models in

68 Ibid., 15.
69 Ibid.
70 Smith, “Curating Architectural 3D CAD Models,” 101.
71 Ball, “Preserving Computer-Aided Design (CAD),” 16.
72 Fallon and Dougherty, “A Pilot Project for Archiving Born-Digital Architecture and Design Data at the Art Institute of Chicago,” 388.

STEP, vendors have been slow to introduce support into their products.[73] The implications of this will vary, but for some projects such as Greg Lynn's Embryological House, "where the form and its means of generation are profoundly related … losing this information may be tantamount to the loss of the project."[74] Even in less extreme cases, it is likely that migration will preclude potential insights into designer intent that might have been gleaned from architect-supplied parameters.[75]

To date, there are no fully-featured automated tools for converting 3D CAD models from one file format to another. As a result, migration 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."[76] For archivists who have acquired these skills or can pull upon the expertise of trained architects, calculation of "validation properties" such as calculated volumes and weights for solids and/or point clouds in a CAD model before and after migration can be useful to ensure that the significant properties of a CAD model survive format migration and interpretation by non-native systems.[77]

3.1.6 Arrangement and description

The Generate Descriptive Information function of the OAIS model refers in significant part to the creation of "that information which is used to discover which package has the Content Information" (the digital object and any associated representation information needed to make it understandable) that is of interest to an end user.[78] In practical terms to an archive,

---

73 Smith, "Curating Architectural 3D CAD Models," 101.
74 Shubert, "Preserving Digital Archives at the Canadian Centre for Architecture," 261.
75 Smith, "Curating Architectural 3D CAD Models," 101.
76 Ibid., 102.
77 Ball, "Preserving Computer-Aided Design (CAD)," 26.
78 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 2-7.

this Descriptive Information takes the form of descriptive metadata—structured information that describes an object—and its relation to related records, i.e., its arrangement. In archival theory and practice, arrangement is seen as its own source of important contextual information, as the profession's foundational theories of*respect des fonds* and original order attests.

Architectural archives have a long history of arranging architectural design records. At the highest level of arrangement, records are assigned to record groups or collections consisting of all of the records created, maintained, and/or otherwise brought together as an integrated unit by a particular architect or firm. Retaining records according to their provenance in this way “ensures that evidence concerning events and processes remains complete and undisturbed within the body of documentation and that the origin and source of each document is completely clear."[79] At the next level down, records are organized in records series, “a defined group of records based on a file system or maintained as a unit because the records result from the same function or activity, have a particular form, or have some relationship based on their creation.”[80] In most cases, project files will form “the fundamental unit for arrangement and description of design and construction records.”[81] Best practice dictates that archivists intellectually arrange and describe all records pertaining to a project together, regardless of their format or physical arrangement.[82]

Within project files, best practice for paper architectural records calls for identification

79 Maygene Daniels, “Arrangement of Architectural Records,” in *A Guide to the Archival Care of Architectural Records, 19th-20th* Centuries (Paris: International Council on Archives, 2000): 68.
80 Kathleen Roe, *Arranging and Describing Archives and Manuscripts* (Chicago: Society of American Archivists, 2005): 61.
81 Waverly Lowell, “Arrangement and Description,” in *Architectural Records: Managing Design and Construction Records*, eds. Waverly Lowell and Tawny Ryan Nelb (Chicago: Society of American Archivists, 2006): 95.
82 Ibid., 97.

of various types of drawings, including elevations, sections, plans, details, and perspectives, at various stages along a project's evolution, including “survey drawing; preliminary design; design (including alternative and variant design); working drawing; contract drawing; presentation drawing; and as-built or record drawing.”[83] It is at this level that CAD models diverge from their analog antecedents. Although it may be possible to identify subsequent versions of a design over a series of CAD files, the ease of saving over previous data introduces the possibility that some steps along the way may be lost to future researchers. Much like an author editing a “manuscript” in modern word processing software, retaining multiple versions of a CAD project as it evolves over time requires conscious effort on the part of the creator. Further, because 3D CAD models in particular are not closed configurations but rather data stores from which forms like sections, elevations, and plans can be generated on request, the arrangement of these files within a record series requires some divergence from traditional practice.

Despite these differences, there is little reason to think traditional archival standards for arrangement and description, including content standards such as the United States' *Describing Archives: A Content Standard (DACS)*, Canada's *Rules for Archival Description (RAD)*, and the *General International Standard Archival Description (ISAD(G))*; structure standards such as *Encoded Archival Description (EAD)* and *Encoded Archival Context—Corporate Bodies, Persons, and Families (EAC-CPF)*; and data value standards such as the *Library of Congress Subject Headings (LCSH)* and the Getty's *Art & Architecture Thesaurus (AAT)*, cannot be applied to born-digital architectural records. When item-level descriptions of CAD files are desired, alternative standards such as the *Cataloging Cultural Objects (CCO)*

---

83 Eleanor Gawne, “Cataloguing Architectural Drawings,” *Journal of the Society of Archivists* 24, no. 2 (2003): 178.

content standard and *Categories for the Description of Works of Art (CDWA)*, which were designed with the unique qualities of visual materials in mind, can be applied to capture descriptive information and stored within a Metadata Encoding and Transmission Standard (METS) record containing structural and technical metadata. Much like audio/visual materials, Word processing documents, and other non-paper records preserved and made accessible by archival repositories, as products and evidence of the day-to-day operations of records creators, CAD and BIM files can be properly intellectually arranged and described alongside archival materials of other formats without significantly breaking from traditional practices.

When a collection is comprised of the records of a single project, an interesting alternative or supplement to the traditional finding aid model for arrangement and description of CAD models is the Project Information Model (PIM) developed by MIT during the FACADE project. A PIM can be thought of as similar to a Building Information Model (BIM) with a wider scope, in that it directly connects a 3D CAD model to related material such as correspondence, in this case through use of an RDF ontology. The development of the PIM was spurred by a realization by MIT that "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."[84] Although other archives may not choose to emulate the FACADE model precisely, this attention to pulling together related documents seems both useful for researchers and very much in tune with archival theory and practice.

One descriptive element unique to the OAIS model is Representation Information.

84 Smith, "Curating Architectural 3D CAD Models," 103.

OAIS defines Representation Information as "[t]he information that maps a Data Object into more meaningful concepts."[85] In general terms, this means that Representation Information provides the user with information necessary to decode or understand the digital object. Representation Information for a 3D CAD model might thus include a file format definition, the software system which natively created and rendered the model, and information about how different software was used within a particular project.. Archivists must be careful to document such information, as what is common knowledge today may not remain so over long periods of time.

3.1.7 Result

The end product of the ingest service is the generation of one or several Archival Information Packages (AIPs) from the data originally accessioned by the archive (Submission Information Packages, or SIPs).[86] The steps above, including normalization, description, and the generation of checksums to verify fixity, ensure that these AIPs "conform to the Archive's data formatting standards and documentation standards."[87] It is these AIPs that will be stored in the archive's data store, and that will be called upon to provide or generate Dissemination Information Packages (DIPs) by the Access service upon request by end users.

3.2 Archival Storage

3.2.1 OAIS definition

Within the OAIS Model, the Archival Storage Functional Entity describes the "one or

85 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 1-14.
86 Ibid., 4-6.
87 Ibid.

more mechanisms, local or remote, for storing digitally encoded information."[88] Its processes include Receive Data, Manage Storage Hierarchy, Replace Media, Error Checking, Disaster Recovery, and Provide Data. A robust Archival Storage service will ensure that data remains accessible to the archive and its users over the long term, and thus plays a central role in the archive's preservation services.

3.2.2 Storage media

Digital data, like all information, must be inscribed or otherwise recorded on some type of physical media, a simple fact which can be easily forgotten in the era of cloud computing and "invisible" interfaces, where computers "present a premeditated material environment built and engineered to propagate an illusion of immateriality."[89] An archival storage medium would ideally be one that would reliably keep data unchanged and accessible to users over long periods of time. Unfortunately, as David Rosenthal and others have pointed out, no such medium exists: "All storage media must be expected to degrade with time, causing irrecoverable bit errors, and to be subject to sudden catastrophic irrecoverable loss of bulk data such as disk crashes or loss of off-line media."[90] Data stored in an archive also face additional threats, including hardware failure; software bugs; communication errors; media, hardware, and software obsolescence; operator error; natural error; and malicious attacks, each of which threaten to destroy or alter an archive's AIPs.[91]

Given the range of potential threats, storage systems utilized by digital archives must

---

88 Ibid., 4-7.
89 Matthew Kirschenbaum, *Mechanisms: New Media and the Forensics Imagination* (Boston: MIT Press, 2008): 135.
90 David Rosenthal, et al., "Requirements for Digital Preservation Systems: A Bottom-Up Approach," *D-Lib Magazine* 11, no. 11 (2005).
91 Ibid.

be redundant and robust. In addition to ensuring that the storage system can tolerate the failure of any of its individual components, “[m]edia, software, and hardware must flow through the system over time as they fail or become obsolete, and are replaced.”[92] Data duplication, whether achieved locally via a RAID array and off-site back-ups, through commercially available cloud services, or over distributed networks such as LOCKSS[93], is a crucial part of protecting against such single points of failure.

Whatever the solution, archives must be aware that storage is not a one-time cost. As Rosenthal stressed in his talk at the Fall 2014 Coalition for Networked Information (CNI), “Preservation and dissemination costs continue for the life of the data, for 'ever'.”[94] Since by all indications the falling cost of disk storage has no longer been exponential since 2010 (breaking “Kryder's Law”) and cloud storage does not appear to lead to significant cost savings for digital archives at scale, the economics of storage media must be carefully considered and budgeted for in order to ensure long-term preservation and access of born-digital architectural records.[95]

Finally, industry standards for “preservation” media change quickly. Not twenty years ago, William Mitchell said that compact disks “have a much longer life and could be thought of as an archival medium,” reflecting a broader consensus of the time.[96] It is possible and even likely that archivists in another twenty years will regard today's writing encouraging the use of hard disks, solid state drives, and “LTO” magnetic tape as digital archives storage media with

---

92 Ibid.

93 David Rosenthal and Daniel Vargas, “Distributed Digital Preservation in the Cloud,” *International Journal of Digital Curation* 8, no. 1 (2013).

94 David Rosenthal, “Talk at Fall CNI: Improving the Odds of Preservation,” *DSHR's Blog*, December 9, 2014, http://blog.dshr.org/2014/12/talk-at-fall-cni.html

95 Ibid.

96 Mitchell, “Architectural Archives in the Digital Era,” 203.

the same incredulity we now experience reading statements like Mitchell's from the 1990s.

3.3.3 Fixity/error checking

Like their analog equivalents, “[m]ost data items in digital preservation systems are, by their archival nature, rarely accessed by users.”[97] In order to ensure that AIPs stored in the archive remain unaltered, an archive must run fixity audits of its holdings at regular intervals. These audits work by comparing a computed cryptographic hash (checksum) to the hash value that was recorded during the Ingest service. A match between the computed and recorded values indicates that the AIP is unchanged. In the event that the values differ, the authenticity of the AIP has been compromised and an uncorrupted version of the original data must be restored from one of the system backups.

3.3.4 Digital repository systems

Archives may choose to take advantage of one of the many digital repository software solutions now available in order to manage the preservation of their digital objects. Many of these services, including DSpace, Fedora Commons, and Fedora-based projects such as Hydra and Islandora, are available as free open source software. Although these services are free to install and implement, they vary in ease of installation and customization and may require service contracts or substantial in-house IT support to run in a production environment.[98]

---

97 Rosenthal, “Requirements for Digital Prservation Systems.”

98 Michel Castagné, “Institutional repository software comparison: Dspace, Eprints, Digital Commons, Islandora and Hydra” (University of British Columbia School of Library, Archival and Information Studies, 2013), http://circle.ubc.ca/handle/2429/44812

Another digital preservation service option gaining traction in the archival community is Archivematica. Like DSpace and Fedora, Archivematica is free open-source software. Unlike the other options, Archivematica utilizes a micro-service design, using “a combination of Archivematica Python scripts and one or more of the free, open-source software tools bundled in the Archivematica system” to create and verify Information Packages, manage storage locations, and perform other tasks required of an OAIS-type archive.[99] As of February 2015, Archivematica also offers ArchivesDirect, a “hosted service offered by DuraSpace for creating standards-based digital preservation content packages that are archived in secure long-term storage” using Amazon S3 and Glacier cloud services.[100] The suitability of such services for archival storage should be closely examined. Nonetheless, ArchivesDirect and similar services such as Preservica offer to greatly simplify the technological commitment an institution must make to engage in digital preservation.

3.3 Data Management

3.1.1 OAIS definition

The Data Management Functional Entity consists of the functions Administer Database, Perform Queries, Generate Report, and Receive Database Updates, that together store and manage the archive's Descriptive Information and system information.[101] Proper management of this information is crucial to ensure that Information Packages can be discovered, accessed, and understood by archive staff and end users alike.

99 “Micro-services,” Archivematica wiki, https://www.archivematica.org/wiki/Micro-services
100 “ArchivesDirect,” http://archivesdirect.org/
101 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 4-10.

3.1.2 Data management

The specifics of database administration and metadata management in general will depend heavily on the content management and other systems utilized by the digital archive and its host institution. These systems must be able to manage the descriptive and system information an OAIS-compliant digital archive requires, or else they must be supplemented by additional services.

One recent example of such a service is the Digital Repository for Museum Collections (DRMC). Developed by Artefactual Systems for the Museum of Modern Art (MoMA), the DRMC acts as an interface between MoMA's Archivematica installation and content management system (The Museum System, or TMS), enabling staff to search and browse digital object metadata, generate reports, and monitor the fixity of stored AIPs. The DRMC as built contains many features suited specifically for preservation of “time-based media and born-digital artworks,” and as such may prove to be similarly useful for architectural collections containing a number of 3D models, animations, and other visual media.[102]

3.4 Administration

3.4.1 OAIS definition

The Administration Functional Entity describes the higher-level managerial activities required to create and sustain an OAIS-type archive. Its functions are Negotiate Submission Agreement, Manage System Configuration, Archival Information Update, Physical Access Control, Establish Standards and Policies, Audit Submission, Activate Requests, and

102 Dan Gilean and Jesús García Crespo, “Introducing the Digital Repository for Museum Collections (DRMC)” (presentation, Code4LibBC 2014, Vancouver, British Columbia, Canada, November 27, 2014), http://www.slideshare.net/accesstomemory/introducing-the-drmc

Customer Service.[103] Through the establishment and enforcement of standards and policies, Administration ensures the long-term viability of the digital archive as a whole.

3.4.2 Establishing standards and policies

Nancy McGovern has famously likened digital preservation to a “three-legged stool,” with each leg representing a different core aspect of a digital preservation enterprise: organizational infrastructure, technical infrastructure, and requisite resources. The first of the legs, organizational infrastructure, “determines the 'what' of digital preservation—the mandate, the scope, the objectives, the staffing of an organization—for engaging in digital preservation.”[104] The establishment of clear and referenceable policies allows an organization to “articulate and institutionalize its commitment,” providing clear direction and purpose to its activities.[105]

Standards exist to help digital archives develop and maintain policies in line with digital preservation best practices, including the Audit and Certification of Trustworthy Digital Repositories (ISO 16363). The *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, published in 2007 by the Online Computer Library Center (OCLC) and Center for Research Libraries (CRL), is helpful as a tool for “objective evaluation … of local capabilities against a set of core criteria for a trusted digital repository” in all three of the areas identified by McGovern.[106]

---

103 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 4-11 - 4-13.
104 Nancy McGovern, “A Digital Decade: Where Have We Been and Where Are We Going in Digital Preservation?” *RLG DigiNews* 11, no. 1 (2007).
105 Daniel Noonan, “Digital Preservation Policy Framework: A Case Study,” *Educause Review*, July 18, 2014, http://www.educause.edu/ero/article/digital-preservation-policy-framework-case-study
106 *Trustworthy Repositories Audit & Certification: Criteria and Checklist* version 1.0 (OCLC and CRL, 2007): 5.

3.4.3 Defining the Designated Community

One of the most important parts of the digital preservation policy for an OAIS-type archive is the definition of its designated communities. In the OAIS model, a designated community is an "identified group of potential Consumers who should be able to understand a particular set of information."[107] Because designated communities constitute the archive's intended user base, their definition is at the heart of the question of significant properties and has implications for description, migration, and access policies. Architectural historians, property managers, urban planners, media studies theorists, visual artists, students, and the original designer of a project may all wish to access archived CAD models for vastly different reasons. As such, an archive preserving these files must think carefully about who its designated communities are and what aspects of these models will be most useful to them.

3.5 Preservation Planning

3.5.1 OAIS definition

The Preservation Planning Functional Entity refers to the processes by which an OAIS-type archive develops and updates its preservation workflows, and includes the functions Monitor Designated Community, Monitor Technology, Develop Preservation Strategies and Standards, and Develop Packaging Designs and Migration Plans.[108] Through active monitoring of the digital preservation landscape and active ongoing evaluation of local strategies and standards, preservation planning ensures that the archive's preservation activities remain effective over time.

---

107 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 1-11.
108 Ibid., 4-14 – 4-15.

3.5.2 Monitor Designated Community

Because preservation priorities and activities in an OAIS-type archive are highly contingent on the archive's designated communities, it is crucial that the archive maintain contact with its end users and evaluate their needs regarding “data formats, media choices, preferences for software packages, new computing platforms, and mechanisms for communicating with the Archive.”[109] In addition, any archive preserving born-digital architectural design records must also remain aware of their users' collective knowledge base and skill sets, including familiarity with available CAD and BIM software, in order to ensure that access to design records is provided in an environment conducive to research. The OAIS reference model suggests several mechanisms for soliciting such information, including surveys, periodic formal reviews, community workshops, and interactions with individual users.[110]

3.5.3 Monitor technology

Because of the quickly-evolving CAD/BIM market, architectural archives must also be sure to remain abreast of technological developments in the field, including the release of new design products, changes to file format specifications, and developments in standards such as STEP and IFC. In order to provide access to the widest range of files possible, architectural archives should also seek to retain licenses for all versions of software the archive can realistically maintain corresponding to actively accessioned and preserved file

109 Ibid.
110 Ibid.

formats, from the most recent release back to the earliest version with a commercially available license. Open source renderers and commercial software with non-time-limited licenses should be preserved for the same reason. Because such software packages are digital objects in their own right, susceptible to threats such as bit rot and obsolescence, best practice may entail storing such programs as their own AIPs within the archive (for additional information regarding this issue, see Section 3.6.3).[111]

Architectural archives would be wise to also monitor developments in related fields, including the defense and aerospace, automotive, and engineering communities. A consortium of aerospace and defense companies from the United States and Europe been actively developing their own standard for long-term preservation and access of 3D CAD models and product data, called Long Term Archiving and Retrieval (LOTAR). LOTAR "draws heavily from existing models and best practices" such as OAIS and STEP, with a "heavy emphasis on checking data against quality and validation criteria … to ensure the data are likely to remain useful … [and] that key characteristics of the data survive when opened in a new system."[112] Tools and practices utilized for this purpose may prove to be immensely helpful to architectural archivists, who face a very similar set of challenges.

3.5.4 Advocacy

As described in Section 2.2.1 of this paper, many of the technical aspects of CAD files and systems that complicate preservation efforts exist because there is little market incentive at present for CAD vendors to address these issues. Solving this problem in a systematic

---

111 Julie Doyle, Herna Viktor, and Eric Paquet, "Long-term digital preservation: preserving authenticity and usability of 3-D data," *International Journal on Digital Libraries* 10, no. 1 (2009): 36.

112 Ball, "Preserving Computer-Aided Design (CAD)," 16-17.

fashion will require a great deal of advocacy for standard formats and high-quality migration functionality between CAD formats. To be successful, this advocacy will likely need to "emphasize the business benefits of a reliable and usable archive of CAD models and the efficiency savings afforded by systems interoperating through common information and data formats."[113] In addition to standards compatibility and robust migration capabilities, archivists should pressure CAD vendors to retain and provide copies of out-of-date software and provide open-ended preservation licenses for non-profit cultural organizations collecting architectural design files. Given reports that Autodesk "may not even have a complete archive of its own software," there appears to be little time to waste on this front.[114]

3.6 Access

3.6.1 OAIS definition

The Access Functional Entity refers to the process by which Dissemination Information Packages (DIPs) are generated and provided to end users in an OAIS-type archive. It contains the functions Coordinate Access Activities, Generate DIP, and Deliver Response.[115] Given the complexity of born-digital architectural design records such as CAD and BIM files, the question of access must be given careful consideration by archives collecting these types of digital objects.

3.6.2 Migration

In an ideal case, normalization of CAD and BIM models to standard formats during the

---

113 Ball, "Preserving Computer-Aided Design (CAD)," 28.
114 Gerald Beasley and Annemarie van Roessel, "Digital Architectural Archives in North America," in *Architecture and Digital Archives* (Gollion: Infolio, 2008): 281.
115 *Reference Model for an Open Archival Information System (OAIS) Magenta Book*, 4-16 – 4-17.

Ingest service should enable end users to view and manipulate copies of these files within a range of software systems. The use of annotations and validation properties, originally recorded in the file's native CAD system, allows users to test the integrity of the models in various systems and choose the reader that exhibits the greatest degree of functionality and least information loss.[116] When users do not require access to the full 3D model, lightweight and visualization formats such as PDF and IGES files may suffice. Ultimately, which migrated versions of a file are suitable will depend upon the archive's designated communities and whether its users wish to merely reference design files, reuse their data, or recover design rationales.

This migration approach, whereby normalized files are accessed on modern software and hardware systems, is most likely to be successful for newer CAD models whose native software suites feature robust export features for standard formats like STEP. Nevertheless, some information loss is always to be expected, particularly when designs were created "on software specifically chosen for its ability to generate and manipulate complex geometries."[117]

3.6.3 Emulation

As an alternative to migration, digital archives may focus their access efforts on providing exact copies of original data in an emulated version of their original environment. As a minimum requirement, this approach would require an archive to collect and preserve copies of obsolete CAD software. Because all software operates within a particular environment comprised of specific operating systems, computer architectures, storage

116 Ball, "Preserving Computer-Aided Design (CAD)," 29.
117 Shubert, "Preserving Digital Archives at the Canadian Centre for Architecture," 261.

devices, input-output devices, and other variables, “[n]ot only the digital object, but also the environment in which it is executed has to be documented and recreated for a faithful reproduction of the rendering process.”[118] This introduces additional layers of complexity into the archive. Julie Doyle, Herna Viktor, and Eric Paquet have suggested a multiple-AIP approach to creating such an environment, where a digital object, an emulator, application software necessary to render the digital object, and metadata associated with the object are stored as four separate AIPs and served to end users together.[119]

Despite the technical challenges associated with emulation, this strategy can potentially offer a number of benefits to an archive and its users:

> Despite the costs involved with emulation, it 'promises predictable, cost-effective preservation of original documents, by means of running their original software under emulation on future computers.' The advantages of emulation are centered on its holistic approach. If one emulates an obsolete operating system, the concerns about loss, authenticity, and error that plague migration largely disappear. And, through relatively new to digital preservation, emulation is a well-established practice in various fields, including engineering and computer science, which provides us with both precedent and guidelines. Furthermore, emulation is a practice that is both reversible, in the sense that original data and programming is nearly always stored as a backup, and verifiable, because one can test and review an emulation immediately after deploying it.[120]

Assuming that the associated technological and legal challenges can be overcome, emulation is the only strategy with the potential to provide loss-less rendering of historical CAD data to users.

Recent advances in emulation technology and services promise to remove many of the

---

118 Mark Guttenbrunner and Andreas Rauber, “A Measurement Framework for Evaluating Emulators for Digital Preservation,” *ACM Transactions on Information Systems* 20, no. 2 (2012): 14:10.
119 Doyle, Viktor, and Paquet, “Long-term digital preservation,” 36.
120 Laura Carroll, et al.,“A Comprehensive Approach to Born-Digital Archives,” *Archivaria* 72 (2011): 78-79.

technical obstacles currently keeping archives from adopting this strategy.[121] The Internet Archive, for instance, has put a great deal of work into in-browser emulation of video game consoles and computers. As of April 2015, a beta version of Emularity—a software program that allows "emulation for anything, *anything*, in the realm of JSMESS, JSMAME, and EM-DOSBOX, which are the three main Javascript emulators running at the Internet Archive"—is available for download on Github free of charge.[122] The bwFLA project at the University of Freiburg has focused on developing Emulation as a Service (EaaS), a "scalable emulation service model" which uses "abstract Emulation Components to standardize deployment and to hide individual emulator complexity."[123] As Dirk von Suchodoletz commented in a 2012 interview in the Library of Congress *Signal* blog, the goal of EaaS is to have emulated environments blend in naturally with existing access services:

> These considerations could lead to a solution which provides seamless access to a variety of different older software: a 1985 home computer a game running in the Multiple Emulator Super System (MESS) emulator; mid-2000 Linux, Windows or Sun Solaris desktops; the mid-90s Apple Macintosh PowerPC architecture and even some modern 3D CAD applications through a single application representing a front-end interface to the emulation services.[124]

Because the EaaS architecture is being designed to be usable by non-experts and "efficient in terms of monetary costs, maintenance and management overhead," it has the potential to remove technological barriers to emulation as an access strategy for digital archives.[125]

---

121 As of 2013, "there only a few memory institutions, such as the National Library and National Archives in the Netherlands, as well as the Royal Danish Library, who are investing in this approach." Dirk von Suchodoletz, Mark Guttenbrunner, and Klaus Rechert, "Report on the First iPres Workshop of Practical Emulation Tools and Strategies," *D-Lib Magazine* 19, no. ¾ (2013).

122 Jason Scott, "Behold the Emularity," *ASCII: Jason Scott's Weblog*, April 14, 2015, http://ascii.textfiles.com/archives/4604

123 bwFLA, "Emulation as a Service," University of Freiburg, http://bw-fla.uni-freiburg.de/eaas.html

124 Bruce Lazorchak, "The Foundations of Emulation as a Service: An Interview with Dirk von Suchodoletz, Part Two," *The Signal: Digital Preservation blog*, December 11, 2012, http://blogs.loc.gov/digitalpreservation/2012/12/the-foundations-of-emulation-as-a-service-an-interview-with-dirk-von-suchodoletz-part-two/

125 Isgandar Valizada, et al., "Cloud Emulation – Efficient and Scaleable Emulation-based Services," *iPRES 2013*,

The other, perhaps more significant barrier to emulation as an access strategy concerns acquiring and preserving obsolete CAD software packages and retaining licenses to run these programs legally. As noted in Section 3.5.4, it would appear that even some CAD software vendors do not have access to older versions of their own software. In the absence of legal deposit requirements for software in the United States, cultural institutions may have to depend on donors and second-hand purchases to begin building software archives. Further, time-limited software licenses have the potential to prevent software from legally being used even if a copy of the software can be found. It bears repeating that this situation can only be solved through effective advocacy, directed at specific software vendors and at a broader legal level. Despite its promise as an access mechanism, “to enable emulation as a generic preservation strategy, today's complex legal issues with regard to copyrights, fair-use exemption, etc. have to be solved, ideally on a supra-national level.”[126]

3.6.4 Copyright

Archives must also ensure that they have documented permission to copy, share, and modify the 3D CAD models and other design records they accession, lest they find themselves in violation of designers' intellectual property rights. These issues should be addressed from the outset and formalized in deeds of gift. The Art Institute of Chicago has addressed this issue with a “Non-Exclusive License for Copyright,” which gives the institution a “license for the life of the copyright to reproduce the digital work for exhibitions, publication,

---

http://purl.pt/24107/1/iPres2013_PDF/Cloudy%20Emulation%20%E2%80%93%20Efficient%20and%20Scaleable%20Emulation-based%20Services.pdf

126 von Suchodoletz, Guttenbrunner, and Rechert, “Report on the First iPres Workshop on Practical Emulation Tools and Strategies.”

and educational purposes … in 'any media now known or not yet invented.'" Such language should be sufficiently broad to cover both access and reformatting, while still allowing creators to "control the copyright for their own purposes."[127]

4 | Conclusion

It is clear that long-term preservation and access of born-digital architectural design records such as CAD files will prove to be a significant challenge for cultural institutions in the coming years. However, the Open Archival Information System reference model provides a very useful framework for thinking through these issues and establishing common ground with groups outside of the profession. Adoption of this model in the form of careful planning, monitoring, advocacy, and action should enable libraries, archives, and museum to rise and meet the challenge of preserving contemporary architectural design for near and long-term use.

127 Fallon and Dougherty, "A Pilot Project for Born-Digital Architecture Data at the Art Institute of Chicago," 381.

5 | Bibliography

Architecture and Digital Archives (Architecture in the digital age: a question of memory), edited by David Peyceré and Florence Wierre, 161-170. Gollion: Infolio, 2008.

Ball, Alex. *Preserving Computer-Aided Design (CAD).* DPC Technology Watch Report 13-02. Digital Preservation Coalition, April 2013. dx.doi.org/10.7207/twr13-02

Barrera-Gomez, Julianna and Ricky Erway. *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house*. Dublin, Ohio: OCLC Research, 2013.

Barrett, Anne R. “Born-digital Architectural Records: Defining the Archivable Record.” Master's paper, University of North Carolina at Chapel Hill, 2012.

Beasley, Gerald and Annemarie van Roessel. “Digital Architectural Archives in North America." In *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*, edited by David Peyceré and Florence Wierre, 277-286. Gollion: Infolio, 2008.

Beetz, Jakob, Stefan Dietze, René Berndt, and Martin Tamke. “Towards the Long-Term Preservation of Building Information Models,” In *Proceedings of the 30th CIB W78 International Conference*, October 9-12, 2013, Beijing, China: 209-217.

Bergin, Michael S. “A Brief History of BIM.” *archdaily.* December 7, 2012. http://www.archdaily.com/302490/a-brief-history-of-bim/.

Bozdoc, Marian. "The History of CAD." *iMB*. 2003. Last accessed April 21, 2015. http://mbinfo.mbdesign.net/CAD-History.htm.

bwFLA. “Emulation as a Service.” University of Freiburg. Last accessed April 21, 2015. http://bw-fla.uni-freiburg.de/eaas.html.

CADAZZ. *CAD Software – history of CAD CAM.* 2004. Last accessed April 21, 2015. https://web.archive.org/web/20150207080928/http://cadazz.com/cad-software-history.htm

Carroll, Laura, Erika Farr, Peter Hornsby, and Ben Ranker. “A Comprehensive Approach to Born-Digital Archives.” *Archivaria* 72 (2011): 61-92.

Clark, Nichola. “The Airbus saga: Crossed wires and a multibillion-euro delay.” *New York Times.* December 11, 2006.

Carlson, Wayne. *A Critical History of Computer Graphics and Animation.* 2003. Last accessed April 21, 2015. https://design.osu.edu/carlson/history/lesson10.html.

Castagné, Michel. 2013. “Institutional repository software comparison: Dspace, Eprints, Digital Commons, Islandora and Hydra.” University of British Columbia School of Library, Archival and Information Studies, 2013. Last accessed April 21, 2015. http://circle.ubc.ca/handle/2429/44812.

Daniels, Maygene. “Arrangement of Architectural Records.” in *Architectural Records: Managing Design and Construction Records*, edited by Waverly Lowell and Tawny Ryan Nelb, 77-90. Chicago: Society of American Archivists, 2006.

Doyle, Julie, Herna Viktor, and Eric Paquet. “Long-term digital preservation: preserving authenticity and usability of 3- D data.” *International Journal on Digital Libraries* 10, no. 1 (2009): 33-47.

“Final Report: FACADE2: MIT and Harvard Collaboration.” Harvard Library Lab. Last accessed April 21, 2015. https://osc.hul.harvard.edu/liblab/sites/default/files/325_final_2013_0.pdf

Fallon, Kristine K. and Carissa Kowalski Dougherty. “A Pilot Project for Archiving Born-Digital Architecture and Design Data at the Art Institute of Chicago." In *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*, edited by David Peyceré and Florence Wierre, 379-392. Gollion: Infolio, 2008.

Franken, Bernhard, Berhold Scharrer, and Inge Wolf. “Save the Bubble.” In *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*, edited by David Peyceré and Florence Wierre, 161-170. Gollion: Infolio, 2008.

Gawne, Eleanor. 2003. “Cataloguing Architectural Drawings.” *Journal of the Society of Archivists* 24, no. 2 (2003): 175-187.

Gilean, Dan and Jesús García Crespo, “Introducing the Digital Repository for Museum Collections (DRMC). Presentation given at Code4LibBC 2014, Vancouver, British Columbia, Canada, November 27, 2014. http://www.slideshare.net/accesstomemory/introducing-the-drmc.

Guttenbrunner, Mark and Andreas Rauber. "A Measurement Framework for Evaluating Emulators for Digital Preservation." *ACM Transactions on Information Systems* 20, no. 2 (2012): 14:1-14:28.

Helfrich, Kurt. "Questions of authenticity: challenges in archiving born-digital design records.” *Art Libraries Journal* 35, no. 3 (2010): 23-29.

John, Elys. *CAD fundamentals for architecture*. London: Laurence King Pub, 2013.

Kirschenbaum, Matthew. *Mechanisms: New Media and the Forensics Imagination*. Boston: MIT Press, 2008.

Kirschenbaum, Matthew, Richard Ovenden, and Gabriela Redwine. *Digital Forensics and Born-Digital Content in Cultural Heritage Institutions*. Council on Library and Information Resources, 2010.

Lavoie, Brian. *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)*. DPC Technology Watch Report 14-02. Digital Preservation Coalition, 2014. http://dx.doi.org/10.7207/twr14-02.

Lazorchak, Butch. "The Foundations of Emulation as a Service: An Interview with Dirk von Suchodoletz, Part Two." *The Signal blog*. December 11, 2012. http://blogs.loc.gov/digitalpreservation/2012/12/the-foundations-of-emulation-as-a-service-an-interview-with-dirk-von-suchodoletz-part-two/.

Lowell, Waverly. "Arrangement and Description." In *Architectural Records: Managing Design and Construction Records*, edited by Waverly Lowell and Tawny Ryan Nelb, 89-106. Chicago: Society of American Archivists, 2006.

McGovern, Nancy Y. "A Digital Decade: Where Have We Been and Where Are We Going in Digital Preservation?" *RLG DigiNews* 11, no. 1 (2007).

"Micro-services." *Archivematica wiki.* Last accessed April 21, 2015. https://www.archivematica.org/wiki/Micro-services.

Misra, Sunitha, Christopher A. Lee and Kam Woods. "File-level Access to Disk Images within a Web Browser," *code4lib Journal* 25 (2014). Last accessed April 21, 2015. http://journal.code4lib.org/articles/9773.

Mitchell, William. "Architectural Archives in the Digital Era." *The American Archivist* 59, no. 2 (1996): 200-204.

Mitchell, William. 1999. "New Digital Technologies in Architecture and their Implications for Architectural Records." In *Blueprints to Bytes: Architectural Records in the Electronic Age*. Boston: Massachusetts Committee for the Preservation of Architectural Records, March 12, 1999: 2-8.

Noonan, Daniel W. 2014. "Digital Preservation Policy Framework: A Case Study," *Educause Review*. July 18, 2014. http://www.educause.edu/ero/article/digital-preservation-policy-framework-case-study.

Olsberg, Nicholas. "Documenting Twentieth-Century Architecture: Crisis and Opportunity." *The American Archivist* 59, no. 2 (1996): 128-135.

Pierce, Katheryn. "Collaborative Efforts to Preserve Born-Digital Architectural Records: A Case Study Documenting Present-Day Practice." *Art Documentation: Journal of the Art Libraries Society of North America* 30, no. 2 (2011): 34-48.

Qi, Jianchang and Vadim Shapiro. “Geometric Interoperability with Epsilon Solidity.” *Journal of Computing and Information Science in Engineering* 6 (2006): 213-220.

Reference Model for an Open Archival Information System (OAIS) Magenta Book. Consultative Committee for Space Data Systems, 2012.

Roe, Kathleen. *Arranging and Describing Archives and Manuscripts*. Chicago: Society of American Archivists, 2005.

Rosenthal, David S.H., and Daniel Vargas. “Distributed Digital Preservation in the Cloud.” *International Journal of Digital Curation* 8, no. 1 (2013): 107-119.

Rosenthal, David S.H., Thomas Robertson, Tom Lipkis, Vicky Reich, and Seth Morabito. “Requirements for Digital Preservation Systems: A Bottom-Up Approach.” *D-Lib Magazine* 11, no. 11 (2005). Last accessed April 21, 2015. doi:10.1045/november2005-rosenthal.

Rosenthal, David S.H. “Talk at Fall CNI: Improving the Odds of Preservation.” *DSHR's Blog.* December 9, 2014. http://blog.dshr.org/2014/12/talk-at-fall-cni.html.

Scott, Jason. "Behold the Emularity." *ASCII: Jason Scott's Weblog*. April 14, 2015. http://ascii.textfiles.com/archives/4604.

Shubert, Howard. “Preserving Digital Archives at the Canadian Centre for Architecture: Greg Lynn's Embryological House." In *Architecture and Digital Archives (Architecture in the digital age: a question of memory)*, edited by David Peyceré and Florence Wierre, 255-264. Gollion: Infolio, 2008.

Smith, MacKenzie. “Curating Architectural 3D CAD Models.” *The International Journal of Digital Curation* 1, no. 2 (2009): 98-106.

Trace, Ciaran. "Beyond the Magic to the Mechanism: Computers, Materiality, and What it Means for Records to Be 'Born Digital'.” *Archivaria* 72 (2011): 5-27.

Trustworthy Repositories Audit & Certification: Criteria and Checklist version 1.0, OCLC and CRL, 2007. Last accessed April 21, 2015. http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.

Valizada, Isgandar, Klaus Rechert, Konrad Meier, Dennis Wehrle, Dirk von Suchodoletz, and Leander Sabel. “Cloud Emulation – Efficient and Scaleable Emulation-based Services.” *iPRES 2013.* Last accessed April 21, 2015. http://purl.pt/24107/1/iPres2013_PDF/Cloudy%20Emulation%20%E2%80%93%20Efficient%20and%20Scaleable%20Emulation-based%20Services.pdf.

von Suchodoletz, Dirk, Mark Guttenbrunner, and Klaus Rechert. "Report on the First iPres Workshop of Practical Emulation Tools and Strategies," *D-Lib Magazine* 19, no. 3⁄4 (2013). Last accessed April 21, 2015. http://www.dlib.org/dlib/march13/vonsuchodoletz/03vonsuchodoletz.html.

Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures." In *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, Ottawa, Ontario, Canada, June 13-17, 2011: 57-66.