**Hybrid Architectural Archives: Creating, Managing and Using Digital Archives**

**Theme: Users: Expectations and Use**

# Defining and Encoding Architectural Information for Digital Archives


## Abstract

An architect's digital files are not immediately useful artifacts in the same way as drawings on paper. A digital file is, rather, an information repository that may be purposed in multiple ways: to convey an idea; to produce one or many technical drawings; to perform structural or energy calculations; to provide a quantity survey; or to calculate conformance to a spatial program. In order to archive this file, the archivist must determine:

1. For what use(s) the architect intended the file
2. Which information contained in the file is of archival interest
3. What "archival" formats exist for encoding that information
4. The process for converting the information of interest from the format in which it currently resides to the desired archival format

Each of these steps is achievable with effort and the appropriate technology tools, formats and methodologies. The technological infrastructure is emerging but not yet mature. The technology tools and formats are not yet set in stone. There is opportunity for informed architectural practitioners, educators, archivists and researchers to define information packages of common interest and define required encodings into standard open formats, extending those formats where required. This effort will bring benefits to both the business of architectural offices and the preservation of cultural heritage.

This paper explores possible expectations and uses of digital archives and the tools and formats available for their creation. It also explores methodologies, recently defined or emerging, for non-ambiguous definition of information content and encoding. Research is drawn from study of digital archives in the museum context as well as research into the business need for information to be exchangeable between firms and tasks (e.g.: architect's design to contractor's cost estimate), not only during design and construction, but also throughout the life of the building to support additions, renovations and maintenance activities.


## Author Bio


Kristine K. Fallon, FAIA
President
Kristine Fallon Associates, Inc.
30 East Adams Street, Suite 1040
Chicago, IL  60603
US
kfallon@kfa-inc.com


Kristine Fallon Associates conducted the original *Collecting, Archiving and Exhibiting Digital Design Data* study for The Art Institute of Chicago, wrote the specification for the DAArch digital archiving software and assisted The Art Institute in developing and testing the software and placing it in the public domain.

President Kristine K. Fallon holds and Master of Architecture degree, is a Fellow of the American Institute of Architects (AIA) and chaired the AIA's Technology in Architectural Practice Advisory Group in 2007. She was lead researcher for the Art Institute project, an advisor to the MIT FACADE project and co-project leader for a research project for the US National Institute of Standards and Technology on information handover in the capital facilities industry.

## Defining and Encoding Architectural Information for Digital Archives

Building projects are leaving less physical evidence of the design process in the form of paper sketches and presentation drawings. More and more, the key images and information that support and document major design decisions are purely digital: building information models, computer renderings, onscreen animations, and analysis data. For over a decade, museums and archives have observed this trend and explored archival techniques for digital design data.

### Background from the Research Community

What exactly should be archived? An architect's digital files are not immediately useful artifacts in the same way as drawings on paper. A digital file is, rather, an information repository that may be purposed in multiple ways: to convey an idea; to produce one or many technical drawings; to perform structural or energy calculations; to provide a quantity survey; or to calculate conformance to a spatial program. In order to archive this file, the archivist must determine:

- For what use(s) the architect intended the file
- Whether the information contained in the file is of archival interest

In the *Collecting, Archiving and Exhibiting Digital Design Data*[1] study (2004) conducted by Kristine Fallon Associates, Inc. for the curatorial Department of Architecture and Design of The Art Institute of Chicago and funded by the Schiff Foundation, the premise was that the archival interest in digital data was analogous to that in physical archives: to document buildings or designers of interest; to capture the moments of intervention by the designer; to establish milestones in the evolution of the design; to save the schemes and options not selected; and to preserve evidence of projects never built and competitions not won. The study provided a critical insight: although designers use many, many digital tools in producing their work, it is possible to identify "outputs"—images or other digital artifacts that the designer chose to communicate to his/her team or client. These outputs included 2D and 3D computer graphics, renderings, photo montages, videos and interactive 3D models. Despite their varied parentage, these outputs could be physically described in a handful of archival data formats. The concept was to preserve what was visible. Just as a paper drawing does not tell us the structural sizing algorithm that generated the column and beam system, this digital archiving approach did not seek to capture underlying information – the attributes, properties or relationships – of the building elements represented in the visible digital object. The advantage of this approach was that it made functional preservation possible – i.e.: the material could be made usable (viewable) over the long term, even though the digital formats and the physical media evolved.

The study suggested a "two-tier" collection. Different preservation strategies were applied to each tier: the outputs were to be functionally preserved, while the source data would be preserved at the bit level. This means that the file would be protected from corruption, but no attempt would be made to ensure its future readability.

Almost immediately there arose in the research community concerns for more robust preservation of the source data. This concern was articulated clearly in Howard Shubert's paper, "Preserving Digital Archives at the CCA: Greg Lynn's Embryological House," presented at the Architecture and Born-Digital Archives conference in Paris in 2007. Shubert rejects preservation based on ongoing migration of digital files for the Embryological House, noting that the architect specifically chose software for its ability to generate and manipulate complex geometries. These geometries are only approximated when converted to archival formats. Therefore, he suggests a dual strategy of preserving the digital files in their original formats and pursuing future emulation of the digital environment in which they were created[2].

All researchers recognize that there are the difficulties inherent in the digital archiving practices within design firms. As firms migrate data across computers over time, file dates are reset and a clear chronology is obscured. Worse, data is frequently deleted – intentionally, through human error or through hardware failure.

Shubert discussed the problem of understanding the contents, chronology and relationships among the many computer files associated with the Embryological House[3]. The *Guidelines to Managing Architectural Record[4]s* published by the European Governance Architecture Urbanism Democracy Interaction program (GAUDI) provides advice for organizing and managing both electronic and paper records from a practical, legal and archival perspective. Nevertheless, archivists and collection managers find great variance in the electronic filing practices of firms.

Similar concerns were also addressed in MIT's US Institute of Museum and Library Services-funded study, Future-Proofing Architectural Computer-Aided Design (FACADE)[5], which is schedule to publish its findings in mid-2009. Research questions for the FACADE Project included:

- What techniques can and should be applied to preserve the native CAD architectural models over archival time frames? Given that CAD models require particular versions of specific software programs to interpret them, is it necessary and sufficient to archive the software as well, or is an "emulation" framework needed for the digital archive platforms that host the material?
- What additional process information is needed to capture the entire building life cycle, and how can that information best be stored in digital archives? Is a new standard necessary for encoding that information, or is a linked document sufficient?
- What other annotations need to be supported to capture the architect's intentions and instructions to the contractors and subcontractors who do the construction (i.e. the Building Information Model) and where and how should that information be kept?

The FACADE researchers reached a similar conclusion to that of the *Collecting, Archiving and Exhibiting Digital Design Data* study – that there are" key files" in the collections: 3D CAD models, 2D CAD drawing sets, client presentations files, and a small selection of other images and documents. These constitute a modest amount of data, typically a few hundred files. These should be given special treatment to insure their easy discoverability and access by all target audiences. The FACADE project extended its target audiences to include teachers looking for exemplar models and the general public,

Two other of FACADE's target audiences – practicing architects and architectural historians – recommended that much additional digital data be kept and made accessible via the archive. This data included all the project correspondence, construction documents, business records, and daily records of the project, constituting tens of thousands of additional records with complex and hard-to-automate relationships. The researchers concluded that processing a collection of that size and creating an accurate relationship map manually would be impossible. They pursued an approach whereby these thousands of second tier files should receive minimal, and automated, metadata to support browsing, and full-text indexing to support keyword search.

The resulting approach to metadata assignment was:

- Semi-automated provision of a common set of metadata elements for every file in the collection, drawn from the US National BIM Standard (NBIMS) project, including: project phase (e.g. competition, design development, construction); file format; document type (e.g. 3D model, plan, elevation, image, document, RFI); zone; and architectural discipline (e.g. structural, mechanical, architecture, plumbing, landscape). The original file system location is kept as a property of each file and is searchable, but is not a browsable facet
- Additional metadata assignment for the key items, as identified by the collection curator
- Full-text indexing of all textual files, searchable via the User Interface
- Ability to browse this metadata to locate items in the collection, and to search for items via keywords and browse large result sets
- Initial display of key items on a timeline organized by building project phase

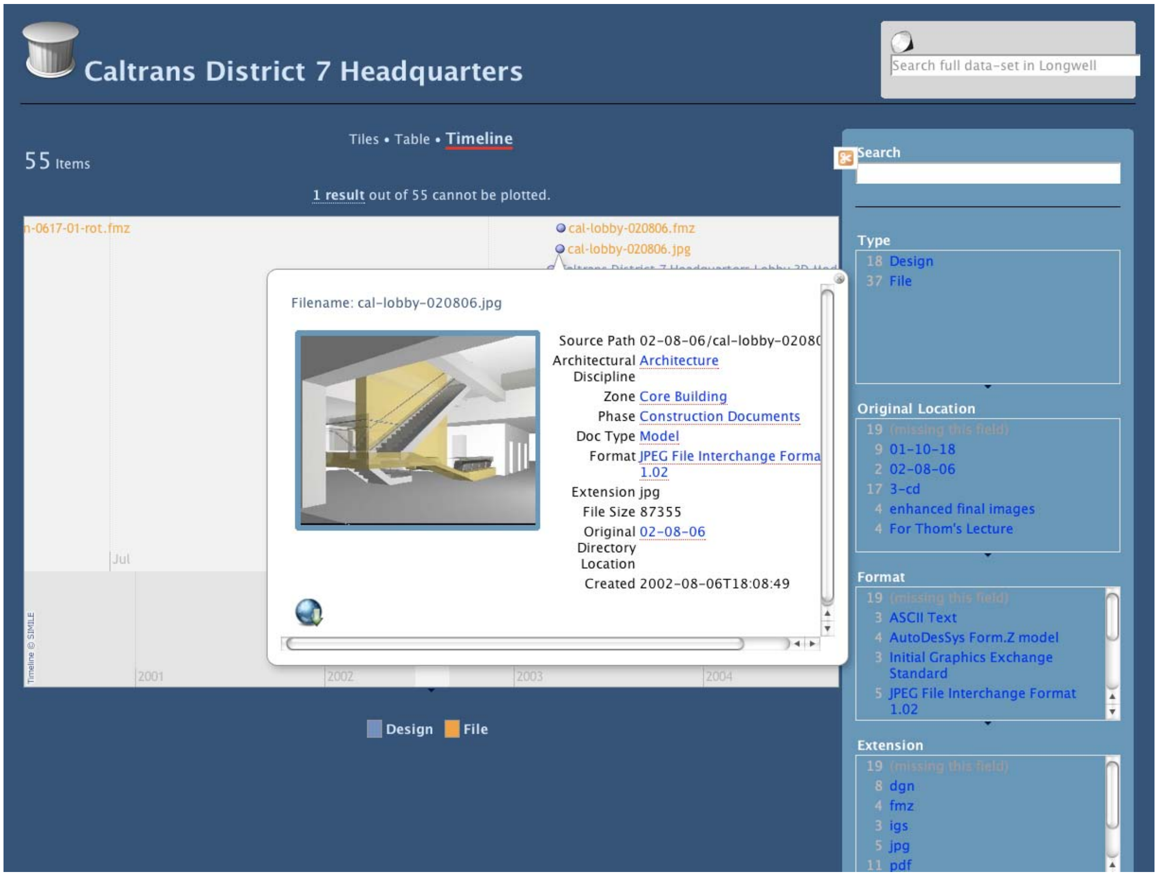


Figure 1: FACADE's prototype user interface
Image courtesy of MacKenzie Smith, Associate Director for Technology, MIT Libraries

Major products of the FACADE project include:

- Recommendation on preservation strategies for 3D models: export to IFC, IGES and 3D PDF in addition to keeping the original format
- Development of a prototype "Curators' Workbench" to assist library and archives staff in processing large collections of digital files rapidly
- Completion of an initial prototype User Interface to architectural data collections in DSpace – a dissemination component
- Development of a Project Information Model (PIM) ontology – a data relationship map for architectural project materials

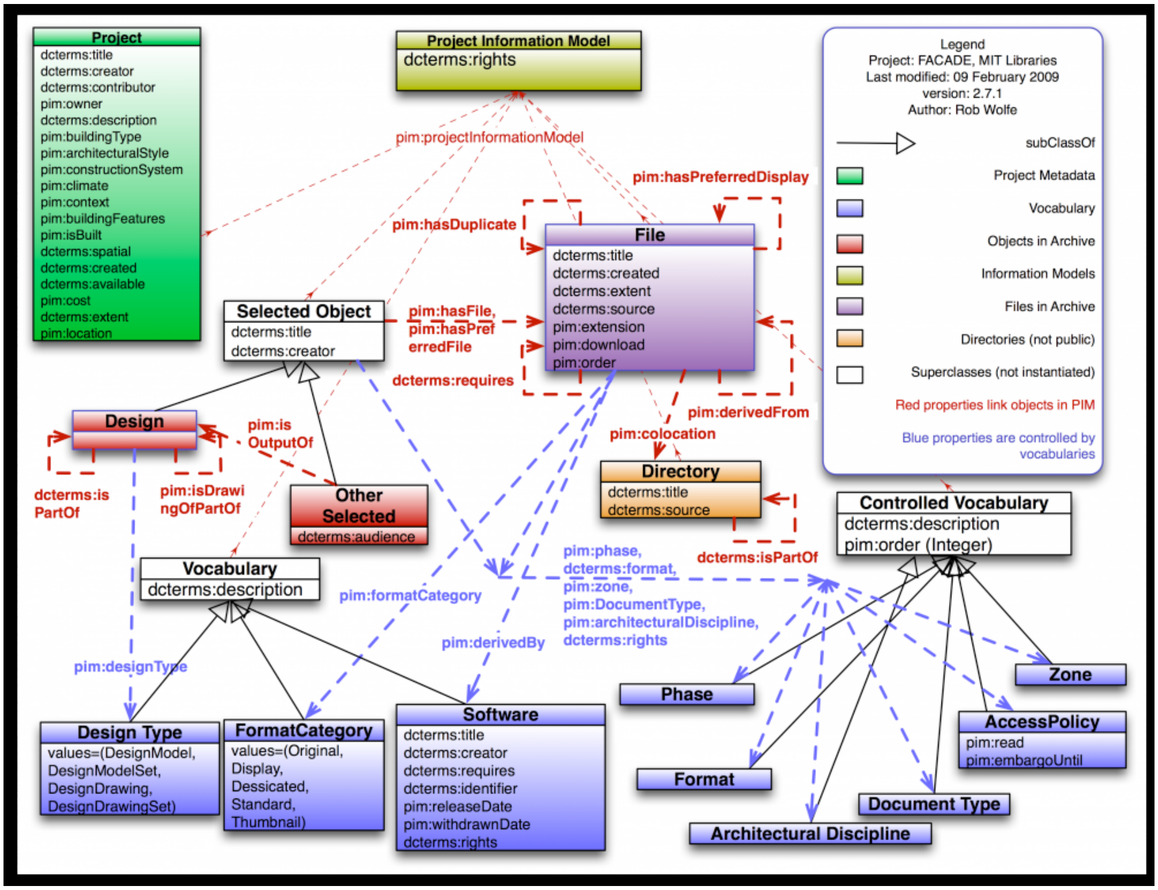


Figure 2: FAÇADE Project Information Model Ontology
Image courtesy of MacKenzie Smith, Associate Director for Technology, MIT Libraries

Also, the FACADE researchers registered information on the file formats they encountered in PRONOM[6]. PRONOM is a service of the National Archives of the UK. It is an online registry of information about file formats, software products and other technical components. Its purpose is to support of long-term access to electronic records and other digital objects of cultural, historical or business value.

The FACADE project also achieved good results with emulation through virtualization. True emulation, which would be the recreation of a specific software product and version within a different hardware and operating system environment, encounters insurmountable intellectual property issues. Virtualization, however, creates a simulated computer environment - a virtual machine - for its guest software. Typically virtualization allows multiple operating systems to run on a single computer or allows a Mac to run Windows creating a virtual PC environment. Virtualization can support the specific computer operating environment needed to run a particular version of a CAD or BIM software product. This would permit simple licensing of the software product versions of interest. What are unknown at this time are the limits of virtualization. There is some level of hardware dependency in all software, both applications and operating systems. Today's market leader in virtualization - VMware®- requires x86 processors. To what extent the hardware physics can be virtualized over archival periods is a question.

Of great interest in this regard is the € 4 million KEEP (Keeping Emulation Environments Portable) project[7], funded as part of the European Union's Seventh Framework Programme (FP7) for research and technological development. Dates are 2007 to 2013.The overall aim of the project is to facilitate universal access to cultural heritage by developing flexible tools for accessing and storing a wide range of digital objects, both static and dynamic: text, sound, and image files; multimedia documents; websites; databases; videogames; and so forth. KEEP will develop an Emulation Access Platform to enable the accurate rendering of these objects, designed for a wide variety of computer systems, to ensure that they remain accessible over the long term. The project is being spearheaded by the Bibliothèque nationale de France, with participants from the Czech Republic, France, Germany, the Netherlands and the United Kingdom.

## Background from Design Practice

The last decade has seen the introduction of Building Information Modeling (BIM) into design practice. BIM changes the representation of a building from drawings to a virtual computer model. BIM creates a single, comprehensive, three-dimensional database of the building. Multiple "views" of this database are possible and include plans, sections, elevations, renderings and quantity surveys. An important distinction between BIM and earlier CAD technology is that BIM's representation of building information is in computer-readable form. Whereas drawings require human interpretation, the computer can interpret a building information model, in terms of its 3D form, its spatial organization, relationship of building elements and properties (e.g. material, color, structural properties, thermal resistance).

The National BIM Standard (NBIMS) project in the United States defines a building information model as "a digital representation of physical and functional characteristics of a facility...and a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition." [8] Much work is underway to create a contractual and business environment that will allow for the discontinuation of drawings, as the project documentation of record, in favor of building information models. Although this change will be gradual, it appears to be inevitable, with far-reaching implications.

If the business of design practice is to require the creation and aggregation of model elements created by multiple parties – architect, structural engineer, mechanical engineer, and so forth – there are clear technical and legal concerns, including:

- Will model elements created by any BIM software product be capable of combining with those created by other BIM software?
- Will a model created in any BIM system be suitable for the full range of analysis applications – structural, daylighting, energy, cost, and so forth?
- How can we trace the origin and modifications made to any model element, for legal purposes?
- How can we know the ownership of the model components, for intellectual property purposes?
- How can the designer designate the status of a model component – provisional or construction-ready?
- How can the various parties query the building model and reliably obtain the information necessary for their work?
- Twenty years from now, when the building requires renovation, will its model be retrievable and usable?

If these questions are answered, as they must be, it will be a great boon to archivists. There are several international standards initiatives that are addressing these concerns.

### Standards Initiatives[9]

Key to the ability to exchange intelligent building information across organizational and system boundaries is standardization of the types, hierarchy and content of the data objects that make up the building model. There must be broad agreement on the information required about a door, a space, a window, a chair, a piece of mechanical equipment and so forth. This is a necessary step for the interoperability of applications using intelligent building models. It permits producers of building products and equipment to create and publish libraries of their components according to a standard framework. Objects from the reference library can then be combined to create any number of truly intelligent BIMs. Each BIM is an assembly of specific components with their properties (height, width, fire rating, and so forth) assigned. Next, software must provide the capability to extract from the BIM the subset of that information required for specific purposes, such as energy analysis, structural analysis, cost estimating, procurement, fabrication, erection and maintenance.

Standard Data Format

The organization working on this is buildingSMART International[10] (formerly the International Alliance for Interoperability). It is a non-profit alliance of building industry participants including: architects, engineers, contractors, building owners and facility managers, building project manufacturers, software vendors, information providers, government agencies, research labs and universities. There are 13 regional chapters: Australasia, Benelux, China, French-speaking Alliance, German-speaking Alliance, Iberian Alliance, Italia, Japan, Korea, Nordic, North America, Singapore, and the United Kingdom and Ireland. Its goal is to develop a universal standard for information sharing and interoperability of intelligent digital building models developed in object-based. The Industry Foundation Classes (IFC) are specifications that define a comprehensive object-based data model for the AEC industry[11]. The IFC core concepts have been endorsed by the ISO as a Publicly Available Specification (PAS) under the ISO label “ISO/ PAS 16739”[12]. They are now in the process of becoming a full ISO standard.

Of particular interest is the Statement of Intention to Support Building Information Modeling With Open Standards[13] signed in January 2008 by the US General Services Administration Public Buildings Service, Senate Properties (the Finnish public real estate enterprise), the Danish Enterprise and Construction Authority and the Norwegian Directorate of Public Construction and Property. The joint commitment was to initiate and participate in open BIM-related research, development, and collaboration efforts. Each Government agency committed to issuing its corresponding BIM requirements, open standard mandates, and adoption schedule. The declared intent was for all major projects to use open BIMs based on IFC on a regular basis but no later than within a two- to four-year ( 2009-2011) timeframe. Signatories also agreed to observe and assess the continuing development of relevant BIM-related standards, such as the International Framework for Dictionaries (IFD) and the emerging Information Delivery Manual (IDM) and Model View Definitions (MVD) approach to defining data exchanges required for the design, construction, and operation of constructed facilities. These are discussed below.

In updating recommendations on archival formats for the Art Institute of Chicago in 2007[14], we added IFC as an archival format for building models in recognition of its increasing popularity and support by software vendors. The FACADE study also recommends IFC as an archival format.

Under development for over ten years, the IFC now provide an extensive language that allows multiple ways to define geometry and to name and measure properties. It is quite possible for two software products to have IFC import/ export capabilities but be unable to exchange information successfully due to different mappings of their native objects to the IFC classes. There are now a number of efforts underway to eliminate this problem.

An established technique is to define specific workflows and use cases for model exchange. An example of a use case would be the handover of building information to the contractor for cost estimating. Each use case may require multiple information exchanges. For each exchange, it is necessary to detail the information to be exchanged and define the entities in the standard format that are required for the exchange. This creates a Model View Definition (MVD). This proven approach derives from the “application protocol” concept that was first used by the Initial Graphic Exchange Specification (IGES) committee in the 1980’s and was a basis for the development of ISO 10303 (STEP) as well as for the development of the Information Delivery Manual (IDM) methodology developed by buildingSMART. The focus of buildingSMART is now less on the development of the data model, although that is periodically updated and extended, and more on the creation of useful IDMs and MVDs.

Classification

Information is organized and classified differently in each building life cycle stage, by different participants and by the various industry sectors. In order for information handed over to be useful, end users must be able to organize, extract and present it flexibly. A good classification framework is critical to managing and providing access to the information.

ISO 12006-2 provides a framework for the classification of information about construction works. Implementations of 12006-2 include Uniclass in the UK and OmniClass™ in the US. OmniClass is described briefly below.

OmniClass Construction Classification System (OCCS)[15] was created through an initiative led by the (US) Construction Specifications Institute (CSI) and buildingSMART. This system provides a set of 15 interrelated tables which classify objects that describe the built environment from a variety of points of view. This classification system was specifically designed to support automated information storage and retrieval, based on a number of ISO initiatives:

- ISO TC59/SC13/WG2 (1988)
- ISO Technical Report 14177 (1994) Organization of information about construction works
- ISO/IS 12006-2 Framework for classification of information
- ISO/PAS 12006-3 Framework for object-oriented information exchange

OmniClass is intended for use in organizing multiple forms of information – electronic and hard copy – in libraries and archives, as well as in classifying project information, including: data exchanges, cost information and specification information. OmniClass creates a series of classification tables based on different ways of organizing information about a building. For example, there are separate tables for phases, design disciplines, building elements, products, and so forth. Each table has a controlled vocabulary and each vocabulary item has a numeric code. The use of numeric coding creates very compact classification tags. OmniClass is a faceted classification system. The tables are designed to work together to provide very detailed and multi-faceted classification of building objects. OmniClass also provides rules for combining the tables. There are 15 OmniClass tables; note that they are non-sequentially numbered:

11. Construction Entities by Function
12. Construction Entities by Form
13. Spaces by Function
14. Spaces by Form
21. Elements
22. Work Results
23. Products
31. Phases
32. Services
33. Disciplines
34. Organizational Roles
35. Process Aids
36. Information
41, Materials
49. Properties

Standardized Terminology

There is a need for common terminology for metadata, including the classification terms in systems such as OmniClass, and object properties. From a commercial perspective, standardizing terminology used for object properties is critical. Consider the importance of properties such as “material” in applications ranging from cost estimating to automated building code checking. The International Framework for Dictionaries (IFD)[16], based on ISO 12006-3, is the leading multi-national initiative. In 2006, the Construction Specifications Institute (US) and Construction Specifications Canada agreed to join with the IFD Partners – STABU (Netherlands) and Standards Norway - to develop a common dictionary of standard terminology for the capital facilities industry. This group is creating a database of terminology and definitions that is multilingual and provides translation capability. It does this by assigning a GUID (Globally Unique ID) to each concept, which can be used to access the proper term in any language supported. The 2x4 IFC release will support IFD information.

## Future Directions

Both The Art Institute of Chicago and the MIT projects resulted in software tools for the archiving of digital design data that are or will be placed in the public domain[17]. Archives, libraries and museums have the opportunity to initiate pilot digital archiving projects, individually or as a collaborative effort, based on these prototypes. Much can be learned through pilot efforts, and that can be fed back to improve the tools.

This community can also look forward to developments from the KEEP project that will permit their collection of digital objects in the formats of the original design software without concern for the loss of functionality and access over time. PRONOM's format registry will support automated identification of the software product associated with each file.

However, there remains the problem of "making sense" of the contents, chronology and relationships among the thousands of electronic files produced by design firms for any project.

This is where the standards work underway on the design practice side holds great interest. Standardization in the areas of building model content, data format, classification, and terminology holds the potential for making the digital objects produced by architects truly and intrinsically discoverable, as well as allowing their utility over the long term. This is important to practitioners, who not only want to see their earlier projects but may need to open and edit the files for a renovation or addition project.

There exist respected metadata schemas and standardized vocabularies in the archival community, such as the Categories for the Description of Works of Art[18] and the Getty's Art and Architecture Thesaurus[19]. However, the pilot digital design data archiving project at The Art Institute of Chicago found that the unique characteristics of both born-digital and architectural materials required developing both new fields and new vocabularies[20]. The classification and dictionary standards under development for design practice should therefore be of interest.

Although OmniClass as a faceted classification system has not yet seen wide commercial use, it is the classification system that has been adopted by the National BIM Standard project in the US. It provides a fruitful approach to relating, searching and examining digital design objects from multiple perspectives. It is intended for application, not only to building model components, but also to documents extracted from or related to the model – drawings, quantity surveys, product and material specifications, requests for information (RFIs), and so forth, allowing for automated associations.

The IFD effort is moving forward quickly, as multiple interested parties, from building product distributors to building code officials, identify the need for standardized and global terminology.

Research findings to date indicate a tremendous stumbling block in gathering digital design data from design practices and preparing it for archiving. The next research project should explore the standards emerging in design practice and discussed here, and seek to harmonize them with those familiar to the archival community. That harmonization will facilitate the collection of digital design archives and enhance the access to and interpretation of such collections.

---

[1] Kristine Fallon Associates, Inc. (2004), *Collecting, Archiving, and Exhibiting Digital Design Data*, CD-ROM, The Art Institute of Chicago, Chicago.

[2] Howard Shubert (2008), "Preserving Digital Archives at the Canadian Centre for Architecture: Greg Lynn's Embryological House," *Architecture and Digital Archives Architecture in the digital age: a question of memory* (Pampelune, Spain: Collection Archigraphie Thématique) p. 261.

[3] *Ibid*, p.262

[4] European Governance Architecture Urbanism Democracy Interaction Program (GAUDI) (2004), *Guidelines to Managing Architectural Records Version 01*, Architectural Archives in Europe, Available at http://www.architecturearchives.net

[5] Information on the FACADE study is taken from the unpublished 6-month FACADE Project interim report: April – September 2008, to the Institute of Museum and Library Services.

[6] http://www.nationalarchives.gov.uk/PRONOM/

[7] http://cordis.europa.eu/fetch?CALLER=FP7_PROJ_EN&ACTION=D&DOC=1&CAT=PROJ&QUERY=011f37a73b31:61ba:091d22f8&RCN=89496

[8] http://www.buildingsmartalliance.org/nbims/about.php

[9] Unless otherwise noted, information in this section is drawn from the NIST/ FIATECH (2007) *General Buildings Handover Guide: Principles, Methodology and Case Studies* (NISTIR 7417), available at http://cic.nist.gov/staff/publications.html (Mark Palmer).

[10] http://www.buildingsmart.com/

[11] http://www.iai-tech.org/products/ifc_specification

[12] http://www.tc184-sc4.org/SC4_Open/Projects/maindisp.cfm

[13] http://www.gsa.gov/Portal/gsa/ep/contentView.do?contentType=GSA_DOCUMENT&contentId=24276&noc=T

[14] Kristine K. Fallon and Carissa Kowalski Dougherty (2008), "A Pilot Project for Born-Digital Architecture Data at the Art Institute of Chicago," *Architecture and Digital Archives Architecture in the digital age: a question of memory* (Pampelune, Spain: Collection Archigraphie Thématique) p. 388.

[15] http://www.omniclass.org/

[16] http://www.iai-tech.org/products/related-specifications/ifd_specification

[17] The Art Institute put DAArch software version 1.0 in the public domain and made it available at http://sourceforge.net/projects/daarch

[18] Baca, Murtha & Harpring, Patricia (eds) (2006), Categories for the Description of Works of Art, J. Paul Getty Trust & College Art Association, Inc., available at http://www.getty.edu/research/conducting_research/standards/cdwa/index.html

[19] Art and Architecture Thesaurus (2000), J. Paul Getty Trust, available at http://www.getty.edu/research/conducting_research/vocabularies/aat/

[20] Fallon and Dougherty (2008), *Op. cit.*, p. 383.