The Art Institute of Chicago
Department of Architecture

Collecting, Archiving and Exhibiting Digital Design Data

Section 2:
Archiving Digital Design Data: Practices and Technology

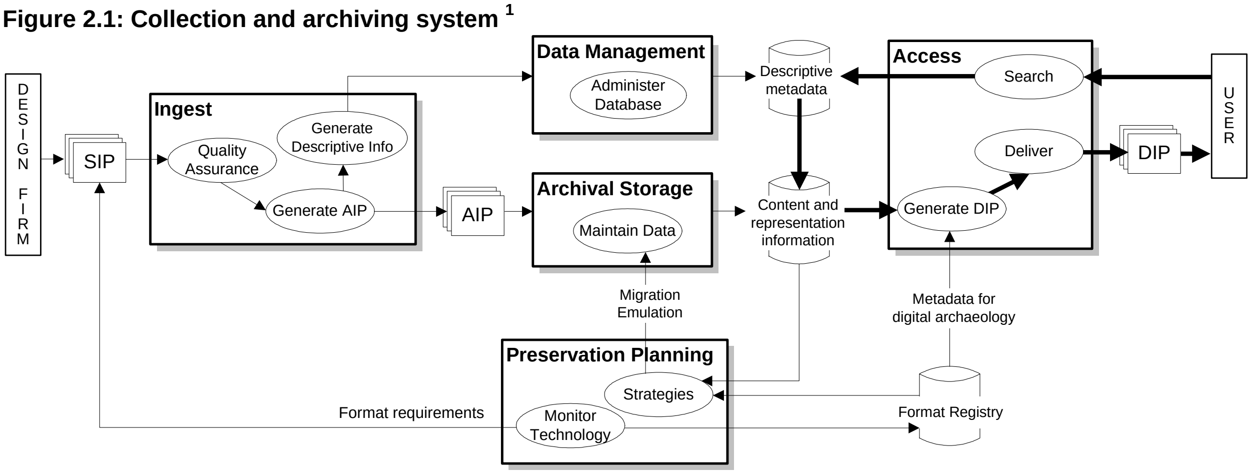
Introduction

This section provides recommendations on practices and technology to be used in archiving and preserving digital design data. It identifies data types and formats to be collected, suggests to design firms practices that will permit institutional archiving of their digital design data, defines methods for cataloging and storing the data, describes tools and methods for accessing and preserving the data and summarizes techniques for digitizing the existing paper-based collection.

There are six distinct stages of the workflow for bringing digital design data from design office to museum archive and for making them accessible to the public. These six stages are: *Preparing, Collecting and Processing, Cataloging, Storing, Preserving* and *Accessing* digital design data. The workflow presented for museum collection and archiving is based on the Open Archival Information System (OAIS) Reference Model for a data repository system. See Figure 2.1. OAIS is an ISO (International Organization for Standardization) standard—ISO 14721:2002—that defines an archival system dedicated to preserving and maintaining access to digital information over the long term.

The recommendations for each stage reflect the collaborative effort of the Advisory Committee for the study, composed of museum curators, archivists, design practitioners, academics, IT managers and representatives from the technology industry, as well as extensive research into precedent archiving and digitization projects, digital data preservation initiatives and CAD viewing and translation technology.

Figure 2.1: Collection and archiving system [1]

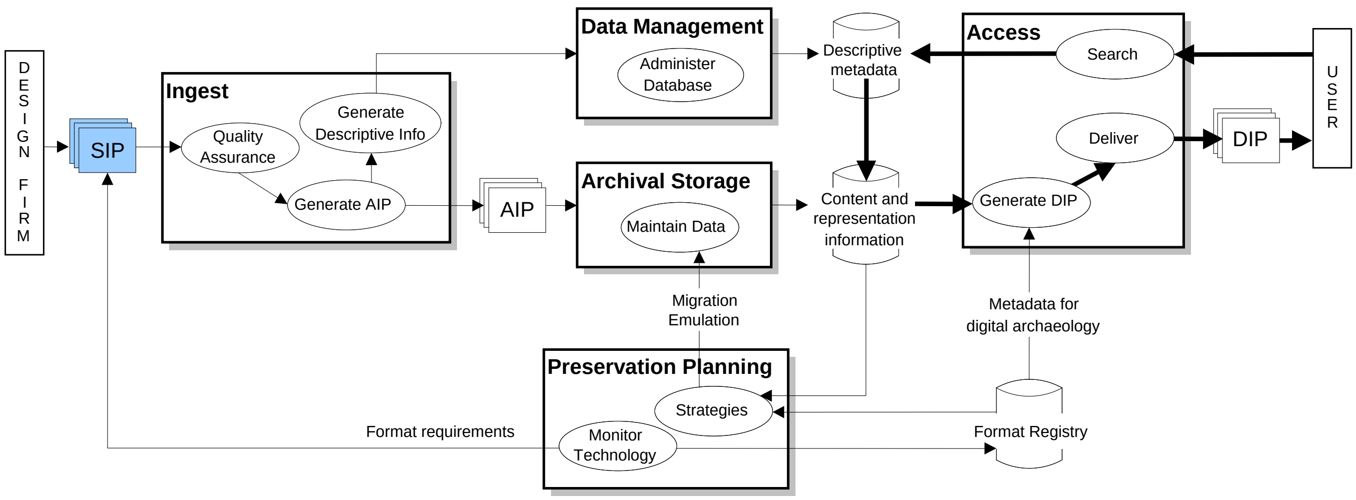


SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

[1] Diagram based on:
Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)* (Washington DC: National Aeronautics and Space Administration, January 2002), publication online, available from ***http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf***; Internet; accessed 29 January 2004.
and
Stephen L. Abrams, "Global Digital Format Registry," *Ready to Wear: Metadata Standards to Suit Your Project, An RLG-CIMI Forum*, 12 May 2003, presentation online, available from ***http://www.rlg.org/events/metadata2003/abrams.ppt***; Internet; accessed 29 January 2004.

Preparing Digital Design Data

Figure 2.1a: Collection and Archiving System: Submission Information Package (SIP)



The first steps in the creation of a successful digital design collection must begin in the designer's office. The design practitioner must organize, name and maintain design data so that a curator or archivist can discern the contents of data files and the time sequence in which they were produced. The designers themselves should preserve important outputs—drawings, images and animations presented to clients—in archival formats. This chapter defines archival format requirements and outlines best practices for design firms to use in organizing and maintaining data.

Submission Information Package (SIP)

Once the Department of Architecture and Design and a design firm have defined the content of a gift of digital design data, the design firm prepares what is known as the Submission Information Package (SIP). SIPs will include the content files and some level of descriptive information including file naming standards or project directory structure for a given set of files. Digital SIPs may be copied onto portable media by the curator or archivist while in the designer's office or may be delivered by the designer via electronic transfer, such as a Department of Architecture and Design FTP server or media such as CD or DVD. The museum will issue a confirmation of receipt to the design office.

Two-Tiered Submission Approach

There are two major types of digital data that designers will submit to a museum collection: output data and native data. The output data represent the designer's intent and include drawings, renderings, animations, photomontages and PowerPoint presentations that document the design process and project milestones. Output data will be formatted in archival formats defined below and thus will be readily viewable. The native data are the source data—the Computer-Aided Design (CAD) files, Building Information Models (BIMs) and so forth—from which outputs are produced. Native data are typically represented in proprietary formats that may require additional tools to access the information.

Archival Format Definition

A museum intends to preserve artifacts forever. This is a unique requirement for preserving digital data. With current technology, this can be achieved for the output data but not for native data. Characteristics of a digital format for output data that will be accessible forever are itemized below:

- Non-encrypted format or published format specification
- Free of patents and other legal restrictions on use
- Independent of specific underlying operating system or hardware functions

- Preserves the appearance and view characteristics of the original
  - Layout
  - Fonts
  - Images
  - Line work
  - Resolution
  - Color
  - Scale
- Contains no externally referenced files
- Broadly used in the archival community
- Readily available viewers.

Archival Formats for Various Digital Content Types

Two well-known archival formats that are appropriate for still images and alphanumeric digital content are a subset of the Portable Document Format (PDF) called PDF/Archive (PDF/A) and Tagged Image File Format (TIFF).These formats are discussed in detail below. The more challenging digital content types that have become prevalent outputs of the architectural process are videos and interactive 3D. Although PDF for Engineering (PDF/E), discussed below, does handle these content types, it is not designed to be an archival format.

The entertainment industry shares an interest in preserving video content and has developed an archival format—MPEG-2—which is discussed in more detail below. Unfortunately, MPEG-2 output is not always supported by the software tools used by designers to create videos. When this is the case, an intermediate format will need to be converted to MPEG-2. When using an intermediate video format, care must be taken to preserve visual quality and so compressed intermediate formats should be avoided. The video should be saved in an uncompressed format, such as uncompressed Audio Video Interleave (AVI). Then a video editing application can be used to convert the uncompressed file to MPEG-2.

If the content does not include audio, another option is available. Animation is, by definition, the rapid display of a series of still images creating the optical illusion of motion. Computer animation content can therefore be defined by its individual still images, known as “frames”, the sequence in which those frames appear and its “frame rate”—the number of frames displayed per second. This understanding provides the mechanism for the long-term archiving of animations. Most computer programs that produce animations will export the frames, creating individual image files numbered sequentially. While uncompressed TIFF would be the preferred format for these images, more frequently supported formats are Joint Photographic Experts Group (JPG), Portable Network Graphics (PNG) and Windows Bitmap (BMP). JPG is not a desirable format because creation of a JPG image usually involves lossy compression, resulting in some degradation of image quality. There are a number of variations of JPG compression, some of which are lossless. Unfortunately, it is often difficult to know which JPG variation animation software is using. For this reason, either PNG, which uses lossless data compression, or BMP, which is an uncompressed format, are preferred over JPG. Both PNG and BMP are discussed briefly below. Frame rate can be specified in a simple ASCII text file accompanying the numbered frames. A video editing application will be needed to reconstitute the images into a video.

Interactive 3D is a recent type of digital design output. The most common use has been to permit someone, typically the client, to navigate around and/or through a proposed design. More recent architectural applications, however, include the ability also to query non-graphic information, such as construction materials, cost codes or item descriptions. Interactive 3D content may be created by a variety of software types, some of which produce purely graphic information and others incorporate non-graphic information. Currently there are a number of proprietary formats commonly used for distribution of interactive 3D content. Two standard formats that support interactive 3D content are Extensible 3D (X3D) and Universal 3D (U3D), both of which are discussed below. Unfortunately, these are not output options that are commonly supported by the programs used by architects, although Adobe Acrobat 3D does use the U3D format for embedding interactive 3D content in PDF files. For outputs from BIM systems, however, there is a very attractive and truly archival format – the International Alliance for Interoperability’s Industry Foundation Classes (IFC). The IFC format is capable of representing a broad range of building information. The IFC format, the preferred archival format for interactive 3D, is also discussed in detail below.

Preferred and Acceptable Formats

This section provides detail on the formats discussed above.

Portable Document Format (PDF)

Portable Document Format (PDF) has become the de facto standard for the exchange of electronic documents and forms around the world. PDF is an open file format specification that preserves document layout and allows embedding of specialized content. The PDF format was created in 1993 by Adobe Systems, Incorporated where it continues to be extended and enhanced. Adobe has placed the PDF format in the public domain. PDF is a publicly available specification, encouraging third party developers to create extensions and tools. Many authoring and viewing applications are available for PDF including Adobe Acrobat and the free Adobe Reader.

In order to maintain visual quality and archival properties when creating PDF documents, it is important to select the correct settings. See *Appendix F: Adobe PDF Settings.*

There are a number of initiatives to create versions of PDF specific to the needs of particular industries and applications.

PDF/Archive (PDF/A)

The PDF/Archive (PDF/A) format is an archival subset of PDF that defines the use of PDF for long-term preservation of and assured access to document content in a consistent and predicable manner. The initiative was begun by the U.S. Courts and spearheaded, beginning in August 2002, by the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES) and the Association for Information and Image Management, International (AIIM International). The international standard (ISO 19005-1:2005) was published in October 2005.

The PDF/A standard trims down the functionality of PDF version 1.4 to include only functions relevant to archival documents. PDF/A documents must be 100% self-contained—all of the information necessary for displaying the document in the same manner as the original must be embedded in the file. Embedded fonts must be free of legal restrictions on embedding and exchange.

PDF/A files must include:

- Embedded fonts
- Device-independent color
- Adobe Extensible Metadata Platform (XMP) metadata—a way of including information about the file within itself.

PDF/A files may not include:

- Encryption
- LZW compression
- Embedded files
- External content references
- PDF transparency
- Multimedia
- JavaScript.

There are two conformance levels in PDF/A: PDF/A-1a (Level A) and PDF/A-1b (Level B). Level B compliance includes what is minimally necessary to ensure the visual appearance of the document. The more stringent Level A compliance includes all the requirements of Level B, but additionally requires that the document's logical structure be included to allow the viewer to view and navigate the document as they could the original.

The PDF/A standard may not be sufficiently comprehensive to archive all forms of design outputs. However, it is the preferred archival format if possible. The focus of the PDF/A standard has been on static documents. Work has already begun on PDF/A Part 2, which is planned to be based on PDF 1.6 and may provide support for additional features such as 3D graphics, audio/video content and JPEG 2000 lossless image compression.

PDF for Engineering (PDF/E)

PDF for Engineering (PDF/E) is an initiative to create an engineering subset of PDF that will define a standard for creating, viewing and printing engineering documents. PDF/E aims to address the need for reliable exchange of engineering documentation. It covers three primary areas—compact, accurate printing of engineering drawings; support of exchanging and managing annotations and comments; and incorporation of complex data into PDF (3D, object data and so forth).

The PDF/E standard is being developed by the Association for Information and Image Management (AIIM) and the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES) along with over 20 organizations participating from both the technical and business side. The first ISO Committee Draft was ratified in May 2006. The final ISO standard is expected in mid-2007.

The PDF/E initiative may provide additional capabilities for capturing design data that will be highly useful to digital design archives.

TIFF

The Tagged Image File Format (TIFF) describes and stores raster image data that comes from scanners, frame grabbers, CAD renderers, photo-retouching programs and so forth. TIFF is able to describe bi-level (two-color only), grayscale and full-color image data in several color spaces and is able to apply a number of compression schemes. TIFF allows the inclusion of special-purpose information such as an embedded color profile, described below under Color Management. It is extensible, meaning that the format is based on a series of tags that can be extended, allowing TIFF to evolve as new needs arise. TIFF is an open and widely supported specification. Version 6.0 can be located on the Adobe Website (***http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf***). The first TIFF specification was published by Aldus Corporation in 1986. Aldus subsequently merged with Adobe Systems Incorporated.

For digitized images, uncompressed TIFF is the archival image format used by the Library of Congress, National Archives and Records Administration and other archival institutions. For born-digital images, such as Photoshop montages or renderings from CAD programs, the recommended archival format is also uncompressed TIFF.

MPEG-2

The MPEG-2 format was initially developed for broadcast television programs, cable and satellite, and has since been adopted for DVD production. MPEG-2 was developed by the Motion Pictures Expert Group (MPEG) in a joint collaborative team with International Telecommunication Union Telecommunication Standardization Sector (ITU-T). It is an international standard (ISO/IEC 13818) and is widely adopted. MPEG-2 is the Library of Congress' preferred format for device-independent digital video for end users and the Library and Archives Canada's preferred format for digital video. Due to its high market penetration and stability, MPEG-2 is the recommended archival format for video data.

Portable Network Graphics (PNG)

Portable Network Graphics (PNG) is a bitmapped graphics format that employs a lossless data compression method. It is an International Standard (ISO/IEC 15948:2004) developed for transferring images on the Internet in a patent-free format. PNG supports palette-based (24-bit) color, grayscale or RGB color modes. It does not support other color spaces such as CMYK but does support embedding International Color Consortium (ICC) color profiles for accurate color matching. Although uncompressed TIFF is the preferred archival format for still images, PNG is an acceptable format for lossless compressed images because it is open, documented, patent-free and supported in many popular graphics applications on multiple operating systems. PNG performs well with images that contain sharp transitions such as text or line art. This makes it an attractive compressed format for capturing individual frames from videos.

Windows Bitmap (BMP)

Windows Bitmap (BMP) is a bitmapped graphics format used by the Microsoft Windows graphics subsystem and as a common raster file format, supported by many image processing programs running under both Windows and other operating systems. It supports black and white, grayscale and up to 32-bit color images. The fact that BMP is commonly supported, well-documented and patent-free makes it an acceptable format for capturing individual animation frames. Although the image file sizes are large, they compress well with lossless compression algorithms such as ZIP, making them a reasonable choice for designers exporting the individual images from their animations. Museums and archives may want to convert BMP images to uncompressed TIFF to simplify long term data management.

International Alliance for Interoperability (IAI) Industry Foundation Classes (IFC)

The International Alliance for Interoperability (IAI) is an organization dedicated to developing a universal standard for information sharing and interoperability of intelligent digital building models. The Industry Foundation Classes (IFCs) define an object-based data model for the Architecture, Engineering and Construction (AEC) industry. They comprise a set of definitions of all the objects encountered in the building industry, and a text-based structure for storing those definitions in a data file. A plain text file is used because that is the only truly universal computer data format. The IFC core concepts have been endorsed by the ISO as a Publicly Available Specification (PAS)—ISO/PAS 16739.

The IFC coverage includes many types of information, including:

- geometry (volume, areas)
- building elements (walls, openings, stairs, doors)
- spaces and spatial structure (space, building storey, building site)
- equipment (ducting, piping, fans)
- furniture (furniture items, furniture systems)
- costing (cost planning, estimates, budget)
- asset identification (maintenance history, inventories)
- associated documents
- work plans (schedules, resource allocation).

The IFC specifications also include support for visualization, such as surface style rendering, materials and lighting specifications.

Many commercial software applications support IFC import and/or export. The IAI has a software certification process, ensuring consistent results. Among the products that have received IFC2x3 Step 2 certification are Autodesk Revit Building, Autodesk AutoCAD Architecture, Bentley Systems Bentley Architecture, Graphisoft ArchiCAD, and Nemetschek ALLPLAN. There are also several IFC viewers available.

Extensible 3D (X3D)

Extensible 3D (X3D) is a royalty free, ISO ratified file format for representing and communicating 3D models. It is known as the Extensible Markup Language (XML)-based successor to the Virtual Reality Markup Language (VRML) format, which is also an ISO standard. X3D is not just a file format for geometry. It supports geometry, lighting, materials, texture mapping, shaders and hardware acceleration. It also supports behavioral modeling and interaction such as animated 3D objects, audio and video mapped into scenes and scripting support.

X3D has a number of different file formats, including XML. X3D is componentized, having a lightweight core and allowing extensibility for various vertical markets through extensions. The core specification is being developed by the X3D Specification Working Group and additional extensions are being developed by domain specific working groups such as the CAD and Medical working groups.

The format specifications are developed by the Web3D Consortium, a group dedicated to creating open standards for the communication of 3D data on the Web and across distributed networks and encouraging the demand for products based on these standards. The group led the development of the VRML 1.0 and 2.0 specifications and today is utilizing its broad-based industry support to develop the X3D specification. Its standardization activities are maintained in close coordination with ISO and the World Wide Web Consortium (W3C).

The abstract specification for X3D was approved by ISO in 2004 (ISO/IEC 19775:2004). The X3D XML and VRML encodings became ISO standards in 2005 (ISO/IEC 19776:2005).

Among X3D’s strengths as an archival format is that it is an open, documented and ISO ratified standard. The availability of XML encoding means that the data can be more easily accessed in the future and has higher potential for integration and support today. X3D's origins from the VRML format, which has endured for a relatively long time, show that it has a good history of support and is likely to persist. Because X3D is not limited to a specific industry, it has a high potential for widespread adoption.

A major drawback to using X3D is the lack of direct support in current digital design tools, making the designer's job of outputting X3D data difficult. At this time, exporters or converters from CAD formats to X3D are rare and direct export from common architectural CAD software is non-existent. Getting data into X3D requires third party tools. One example is PolyTrans from Okino Computer Graphics. PolyTrans is a powerful 3D translation and viewing tool that supports import of various CAD formats and export to X3D and many other formats. The base PolyTrans product supports import of many 3D formats, but generally not those of architectural CAD products.

Universal 3D (U3D)

Universal 3D (U3D) is an open and extensible file format for interactive 3D designs. The format was developed by Intel and the 3D Industry Forum (3DIF) for sharing 3D models on the Internet and in common office applications. U3D is designed as a lightweight format mainly for graphical representation of 3D designs. In order to reduce file size for fast Internet downloading and viewing, U3D strips out most of the non-graphical object data. Although U3D can support some lighting, material and surface information, it doesn't capture object properties such as those produced by BIM applications. The format is primarily intended for visualization.

The 3DIF is a group of technical and corporate users of 3D graphics technology from multiple industries. Participants include Adobe, Bentley Systems, Boeing, HP, Intel and Right Hemisphere. The group is working with Ecma International, an international standards body, to develop the U3D format for submission as an ISO standard.

One of the advantages of the U3D format is that there are readily available tools for outputting content. Ecma lists over a dozen authoring tools including Bentley MicroStation and Adobe Acrobat 3D. Acrobat 3D is notable because it provides the ability to output U3D data from many CAD and BIM applications. Using Acrobat 3D, designers import 3D models from major CAD applications and embed them into PDF files. With the free Adobe Reader version 7 or newer, viewers can view and manipulate the model data. The large install base of Acrobat Reader gives this format a high potential for market penetration.

Acrobat 3D is targeted primarily at the Mechanical Computer-Aided Design (MCAD) industry with the majority of support for MCAD file formats. Architectural CAD file formats that are directly supported include AutoCAD and MicroStation. For applications that do not have direct import support in Acrobat 3D, such as Autodesk Revit, a separate Toolkit utility is provided that allows the capture of 3D geometry displayed in OpenGL mode and converts it into Adobe PDF. Once the model is correctly captured in Acrobat 3D, it can be saved as a PDF with the model data embedded as U3D.

U3D data is encoded in a binary format, which makes it less desirable for archiving than a text format, although this does not preclude it from being a candidate for archival storage. Sustainability of U3D is supported by its being open, documented and widely adopted among this category of data types. U3D has a relationship with the PDF/E format (submitted for ISO ratification) in that the 3D file format specified by PDF/E is U3D. The specification for PDF version 1.6 references U3D. Although the U3D specification is a separate standard and separately maintained, its reference in the PDF specification suggests that support for the format will be continued in future versions PDF.

Best Practices for Design Firms

From the archival institution's perspective, it is desirable for the designer to:

- Identify key outputs that the designer feels represent his/her intent and that document design evolution and milestones
- Pay attention to resolution, compression and color when creating the key outputs
- Preserve, rather than overwrite, these outputs and the native data from which they are produced as the design process moves forward
- Organize and name the digital design data in a way that makes the project milestones and data associations apparent.

Identifying Key Outputs

The designer knows intuitively which drawings, images or animations best capture design intent.

Resolution and Compression

Images that convey design intent and that are intended for use in print publication should be created for a print resolution of 600 dpi. An 11x8.5-inch color image at 600 dpi requires 6600x5100 pixels and is about a 100MB file. See Table 2.1 for relationships between pixels, inches and file size.

Derivative images of lower resolution and file size can be created. For all images created, even low-resolution images intended for electronic viewing, it is good practice to first save an uncompressed version in TIFF before saving derivative images in compressed formats such as JPG. All images submitted should be in uncompressed TIFF format for long-term preservation.

Table 2.1: Relationship between Pixels, Inches and File Size for Images
Note: DVD storage value used is 4.7 GB each

Dimensions in Pixels	Image Dimensions in Inches				File Size (Grayscale)	No./DVD	File Size (Color)	No./DVD
	72 dpi	*200 dpi*	*400 dpi*	*600 dpi*				
400 x 300	5.6 x 4.2	2.0 x 1.5	1.0 x 0.8	0.7 x 0.5	0.120 MB	39,166	0.360 MB	13,055
640 x 480	8.9 x 6.7	3.2 x 2.4	1.6 x 1.2	1.1 x 0.8	0.307 MB	15,299	0.922 MB	5,100
1024 x 768	14 x 11	5.1 x 3.8	2.6 x 1.9	1.7 x 1.3	0.786 MB	5,979	2.36 MB	1,991
1600 x 1200	22 x 17	8.0 x 6.0	4.0 x 3.0	2.7 x 2.0	1.92 MB	2,447	5.76 MB	815
3000 x 2250	42 x 31	15 x 11	7.5 x 5.6	5.0 x 3.8	6.75 MB	696	20.2 MB	232
4400 x 3300	61 x 46	22 x 16	11 x 8.3	7.3 x 5.5	14.5 MB	324	43.6 MB	107
6800 x 4400	94 x 61	34 x 22	17 x 11	11 x 7.0	29.9 MB	157	89.7 MB	52
10,200 x 6600	142 x 92	51 x 33	26 x 17	17 x 11	67.3 MB	69	201 MB	23
19,200 x 14,400	267 x 200	96 x 72	48 x 36	32 x 24	276 MB	16	829 MB	5

Source: Kristine Fallon Associates, Inc.

Additional information on image resolution can be found in the *Digitizing the Existing Collection* chapter.

The characteristics important to the video data type include:

- **Clarity:** Clarity refers to the visual quality of the video. Clarity is first affected by the settings used during the recording process. It can also be greatly affected by the capabilities of the file format in which the animation is saved.
- **Size:** Size refers to the pixel dimensions of the video, given as a number of horizontal and vertical pixels. Larger size is preferable to smaller size, although sizes in excess of the display capabilities of typical output devices (monitors, projectors, etc.) are not advised. At the time of this writing, typical computer projectors support a resolution of 1024x768 and large monitors support resolutions up to 1600x1200. For reference, the NTSC broadcast standard used in the USA, Canada and Japan has a resolution of 640x480 and HDTV broadcasts have a resolution of 1280x720 or 1920x1080.
- **Frame rate:** Frame rate refers to the speed at which the images are shown in succession or more specifically, the number of still images per unit of time. Frame rate is commonly indicated in terms of frames per second (FPS). To achieve a flicker-free animation, a minimum frame rate of 15 FPS should be used. Frame rates in excess of 30 FPS will not provide any perceivable benefit. For reference, the NTSC broadcast standard specifies a frame rate of 29.97 FPS and film is shot at 24 FPS.

Although MPEG-2 is the preferred archival format for video, most CAD, BIM and many 3D modeling applications have very limited options in video export formats. Although these applications list many export formats, these are often proprietary and should be avoided. To save a video in an archival format such as MPEG-2, an intermediate format must often be employed. When performing multiple transformations like this, it is essential that the visual quality be maintained as much as possible. Do not export the video to a compressed file and then convert it to an archival format since multiple encode-decode cycles will degrade the visual quality. If not supported directly by your software application, the preferred method of getting a video into MPEG-2 format, is to export the video to an uncompressed format. Then, use a video editing application to convert the uncompressed file to MPEG-2 format. This method performs only one compression action, which results in better quality.

Color Management

Maintaining color fidelity from the designer's computer to a museum archive and then to exhibition is a challenge. Color management involves careful translation of color values from the source device, such as a designer's monitor, to the destination device, such as the book publisher's printing system. The most difficult aspect of this process is that there is no way of knowing precisely what output method—whether digital or print—will be used to display the content in the future. The best way to ensure color consistency is to follow sound color management techniques within the firm's day-to-day activities.

A color management system (CMS) is a group of software tools and hardware measurement instruments that work together to identify and map the color values of the source device to the output device. The color management process involves three elements: the source and destination profiles, the profile connection space and the color management module.

- The profile tells the CMS the relationship between the red, green and blue (RGB) values of the device—scanner, digital camera or computer monitor—and the corresponding profile connection space values. RGB is an additive color space used by scanners, digital cameras and computer monitors. Cyan, magenta, yellow, black (CMYK) is a subtractive color space used by printers. Profiles can also be abstract working spaces such as sRGB or Adobe RGB (1998). A source profile defines how to convert colors from the first color space (e.g., monitor's color profile) to the profile connection space. A destination profile defines how to convert colors from the profile connection space to the target color space (e.g., printer's color profile).
- The profile connection space (PCS) acts as the go-between that reconciles the RGB or CMYK values of the input device's space and the output device's space. The two standard PCS's chosen by the International Color Consortium (ICC) are CIE-XYZ and CIELAB, two color spaces developed by the Commission Internationale de l'Eclairage (CIE).
- The third element is the Color Management Module (CMM), which is the engine that uses the profiles to convert between source and destination color spaces via the PCS. Software applications such as Adobe products, CorelDRAW and QuarkXPress have color management systems that can be configured based on the needs of the user. Macintosh and Windows operating systems provide their own color management systems: Apple's ColorSync and Microsoft's Windows Color System (formerly Image Color Management, pre-Windows Vista), respectively.

Hardware Calibration

The first step to effective color management involves calibrating each device—computer monitors, printers, scanners and digital cameras—and creating a color profile that describes the way the device handles color. The color profile is typically in an ICC format. While not essential to the digital archiving process, design firms should calibrate monitors and output devices within their office to ensure they are reproducing their on-screen images as accurately as possible.

To calibrate and profile a computer monitor, either hardware devices—"colorimeters" or "spiders"—or software applications can be used. Colorimeters or spiders are devices that are placed on the screen of the monitor and will take red, green and blue readings, white points, black points and gamma level. If the levels are severely off target, they will alert the user that some manual adjustments need to be made. If only small adjustments are needed, the software can make them automatically. Examples of colorimeters are: Pantone ColorVision Spyder line, Integrated Color Solutions basICColor display 4 and GretagMacbeth EyeOne products. Other less sophisticated, and potentially less accurate and less consistent, color calibration software packages rely on the user's eye to match red, green and blue colors provided by their screen to ones presented by the software. An example of a visual calibration tool is Adobe Gamma, which comes standard with Adobe Photoshop.

For the highest level of accuracy in color management, CRT monitors should be calibrated once per week and LCD monitors less frequently. One available software package for calibrating and profiling a scanner or digital camera is GretagMacbeth Profilemaker. This provides a printed color target to be scanned along with a color data file to compare with the scanned target. The software will compare the two images and will build an ICC color profile for the scanning device. A similar process is used for digital camera profiling.

Documenting the Color Source

The second and most important step in color management is embedding the source color profile. TIFF and PDF accommodate embedded color profiles. If the digital content was captured by digital photography or scanning, the capture device's profile should be embedded. With born-digital content or content that has been manipulated after capture in an application such as Photoshop, the working space should be embedded. A working space is a device-independent definition of color. In recent years, as more output has remained digital-only and never printed, device-independent RGB working spaces such as sRGB have become more commonplace. The workflow for creating images and embedding color profiles during the design process should be tailored to the individual design firm and its set of digital design tools. The following are some best practices suggested as a starting point.

Designers should inform themselves of the color management capabilities of digital design tools they use. For AutoCAD users, there is a third-party color management package—M-Color 9 by Motive Systems. If the CAD program itself does not give an option to embed a profile, the image should be assigned the correct profile in a color management tool like Adobe Photoshop or Acrobat. Color Settings in Photoshop should be set so that the program will prompt the user if an image without an embedded color profile is opened. The user is then given the option to assign a profile from a drop-down menu. This will embed the selected profile without changing the color values of the image.

For photomontages in which images from many different sources such as CAD renderings, digital photographs and scanned sketches are being assembled in Photoshop, it is important to choose a large working color space. A working color space in Photoshop is used to map images with different color profiles to a common space and will be embedded in the final image document. To avoid losing color data from digital photos or scanned images whose color spaces have a large color gamut, it is important to choose a large working color space such as Adobe RGB (1998). Preferences on working color spaces are specified in the Color Settings dialog box in Photoshop.

An additional specification made in a color-managed file is called the “rendering intent,” which dictates the way in which the color gamut—the entire range of hues reproducible by a given device—is mapped from the source to the destination device. For example, computer monitors use the additive RGB color space while printers use the subtractive CMYK color space. The color gamut for RGB is different from CMYK and therefore, not all colors can be mapped accurately. There are three primary rendering intents that describe different approaches to mapping: colorimetric, perceptual and saturated. Colorimetric—either absolute or relative—is the strictest approach to mapping and should be used when literal color accuracy is paramount. Colorimetric mapping will find in the smaller color space the “closest possible” match to the color in the larger color space. It is preferred for images such as company logos where it is important to find the closest possible match to a color. Perceptual mapping is a less rigid method that is preferred for photos. It maps based on the *relative* color differences and it may even change colors that can be matched for a better overall look. Saturated rendering intent will map to colors that can be best represented or “most saturated” on the destination device. It is preferred for images such as business graphics or other schematic material where it may be more important to have the best saturated colors than to have an accurate color produced in a poorly rendered way. The designer should identify the desired rendering intent. Perceptual mapping is recommended for renderings and photos and relative colorimetric is recommended for line work.

Mapping Color Values for Output

Once an image has an embedded color profile, software such as Adobe Photoshop performs the function of the color management module to map the color values from the source to destination device.

Pantone

Pantone is a standard for color communication that may aid in the color management process if the system is used by the design firm. Pantone has a numeric representation for hundreds of colors, known as the Pantone Matching System, with specified formulas for mixing inks for print. Photoshop allows designers to select a Pantone color. This system might be applicable if the designer works in Pantone colors and the museum’s publisher uses the Pantone Matching System for inks.

Organizing and Naming Data

There have been Architecture, Engineering and Construction (AEC) industry initiatives to define and publish sound practices for organizing and naming design data. Two are discussed in this section. The *Guidelines to Managing Architectural Records* published by the European Governance Architecture Urbanism Democracy Interaction program (GAUDI) emphasizes the use of metadata to improve access to, and preservation of, digital records. The Construction Specifications Institute’s (CSI) Uniform Drawing System (UDS), incorporated in the

U.S. National CAD Standard (NCS), describes the organization of digital drawing files into standard computer folders and provides naming conventions for both the folders and the files.

Governance Architecture Urbanism Democracy Interaction (GAUDI)

The *Guidelines to Managing Architectural Records* published by the European Governance Architecture Urbanism Democracy Interaction program (GAUDI) is a helpful resource that discusses the importance of records management for design firms. The document provides advice for organizing and managing both electronic and paper records from a practical, legal and archival perspective.

The GAUDI guidelines recommend that firms have formal, written policies for record creation, organization, retention and management. For electronic data, such policies must be put into effect as soon as records are created. An organized system saves the firm considerable time and makes it easier to exchange documents and data with collaborators. Formalizing the policies facilitates adherence and provides a map to the documents for future archivists or records managers. All staff should be aware of, educated on and involved in proper records management.

Records need to be managed throughout the project's lifecycle, but particularly at major milestones. At such milestones, team members may take the opportunity to review what documents they have, purge what is unnecessary or redundant, select what should be preserved, and ensure that all records are properly filed.

The GAUDI guidelines suggest that design firms create a filing system based on the functional sections of their practice—administration, project management, design and so forth—and develop consistent ways of naming projects and phases. However, they provide no specific guidance on this organization and naming.

For electronic records, the GAUDI guidelines recommend the use of metadata to aid preservation and access. The GAUDI workgroup developed a sample metadata element set for describing documents based on the Dublin Core. The Dublin Core Metadata Element Set is a universally recognized set of elements to describe information resources. Dublin Core includes fifteen elements—Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title and Type—and can be extended through the use of qualifiers. Dublin Core is intrinsic to the DSpace repository on which The Art Institute of Chicago's DAArch system is based. Although not nearly as comprehensive as the Categories for the Description of Works of Art (CDWA) metadata scheme, Dublin Core provides a recognized starting point for classifying records. Dublin Core is discussed in depth in the *Cataloging Digital Design Data* chapter.

The GAUDI guidelines provide a sample for describing electronic records using the Dublin Core schema:

- **Title or project name:** A name given to the document
- **Creator:** A person primarily responsible for making the content of the resource
- **Subject or keyword:** A topic of the content of the document
- **Description:** An account of the content of the document
- **Contributor:** A person or persons responsible for making contributions to the content of the document
- **Date:** A date of an event in the lifecycle of the document
- **Type:** The nature or genre of the content of the document
- **Format:** The physical or digital manifestation of the document
- **Rights:** Rights Information about rights held in and over the resource
- **Place of storage:** Information about the place the document is stored.

Assigning metadata can take on many forms, including at a simple level, the location of the file within a folder structure. A more direct application of metadata is to provide basic information about the document in the file's Properties field. Most applications have Properties (or similar) for their file types, where a creator can specify such information as the title, author, dates, keywords and other notes about the document. For example, in Microsoft Word, click File → Properties or in AutoCAD, click File → Drawing Properties and review the many fields of metadata that can be populated. The fields in the document Properties can usually be extended by adding custom fields. To create a comprehensive metadata record for files, creators should add custom Property fields to equal those recommended by GAUDI or the Dublin Core. These same fields can be searched when looking for information. Having metadata within the file facilitates the future archivist's or data manager's job.

The full *Guidelines to Managing Architectural Records* document is available at:

http://www.archivesarchitecture.gaudi-programme.eu/fichiers/t_pdf/14/ pdf_fichier_fr_Prescriptions_en_anglais,_version_web01.pdf

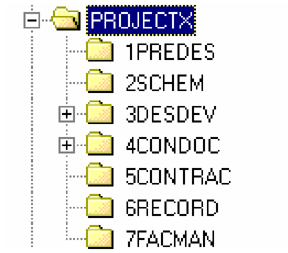
Uniform Drawing System (UDS)/National CAD Standard (NCS)

The Construction Specifications Institute (CSI) has been involved in developing U.S. standards and guidelines for graphic documentation for many years. These efforts culminated in CSI's Uniform Drawing System (UDS), which is a major portion of the U.S. National CAD Standard (NCS). UDS and NCS provide comprehensive guidance for periodic checkpointing, organization and naming of electronic drawing files.

UDS recommends that all project data be copied to an archive folder at major milestones and backed up. It also provides specific recommendations for an organization of project data that corresponds to the major project milestones, with the following subfolders:

- 1PREDES (programming and pre-design phase)
- 2SCHEM (schematic design and concept phase)
- 3DESDEV (design development phase)
- 4CONDOC (construction document phase)
- 5CONTRAC (contract submittal phase)
- 6RECORD (record documentation phase)
- 7FACMAN (facility management phase).

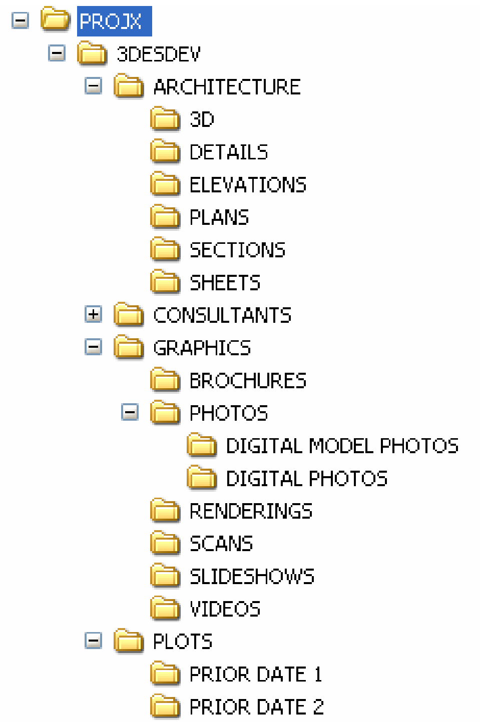
Figure 2.2: ***Uniform Drawing System*** **Project Phases**

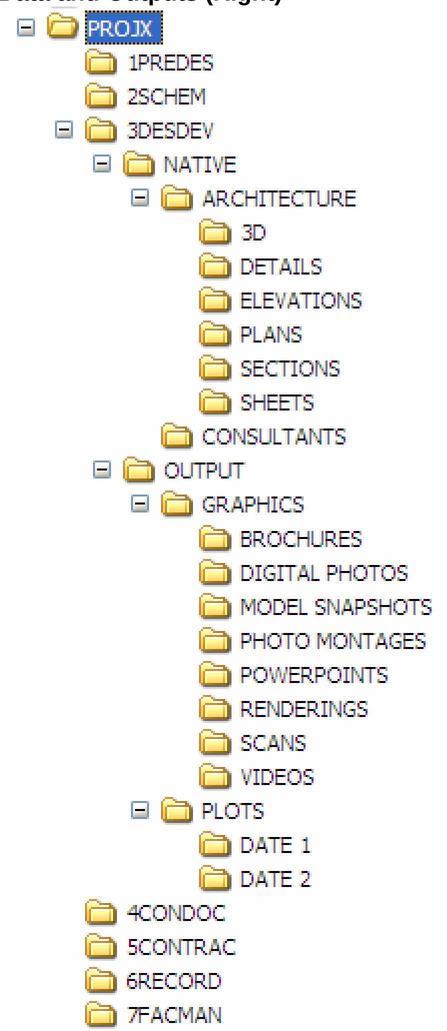


Design firms may choose to organize files by native CAD and output type. The file directory organization used by architecture firm Murphy/Jahn can be seen in Figure 2.3. An integration of the UDS project phase folder names and an output and native format classification can be seen also in Figure 2.3. Consultant's files are included in the folder structure, but will not be part of the Submission Information Package to the museum without prior permission from the consultant firms.

Because many electronic documents, particularly CAD files, have externally referenced files, such as "xrefs" in AutoCAD and image files for materials in Autodesk VIZ, it is important for the design firm to embed all external files into one file before submission to the museum. If it is not possible to embed all externally referenced files, the linkages should be clearly documented. For example, VIZ has a function called by selecting File → Summary Info that will output a text file of all referenced files and their locations. This should be done before moving the files to archive directories.

Figure 2.3: Murphy/Jahn Directory Structure (Left)
Directory Structure with UDS Project Phases, Native Data and Outputs (Right)





The UDS/NCS have established a standard naming schema for native CAD models and sheet files. As design moves toward 3D CAD and intelligent building models (BIM), some of these naming conventions may become obsolete. However, the following file naming schemas are applicable to most of the digital drawings and 3D models produced in architectural practices today.

Note that the UDS/NCS allow an option of including a five-digit project identification prefix in any filename. Use of this option within design firms would be helpful to archiving and long-term data management.

Model Files (Native Data)

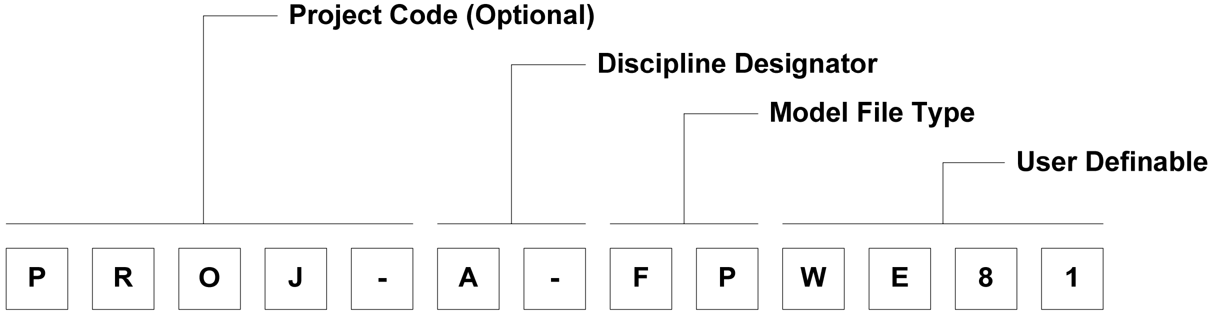
Although these file naming conventions were developed with traditional CAD systems in mind, the approach embraces the concept of a "model", which makes the file naming conventions appropriate for BIM data.

The native CAD model, which contains building geometry and physical components, is named beginning with an optional five-digit project code followed by:

- Discipline designator
- Two-letter model file type
- User-definable field.

See Figure 2.4 for a sample native CAD model filename and Tables 2.2 and 2.3 for a list of *Discipline Designators* and *Model File Types*, as published in the *National CAD Standard, Version 2.0*. [1]

Figure 2.4: CAD Model Naming Schema



Source: Kristine Fallon Associates, Inc.

Table 2.2: Discipline Designators

Discipline Designators	
G	General
H	Hazardous Materials
C	Civil
L	Landscape
A	Architectural
I	Interiors
Q	Equipment
F	Fire Protection
P	Plumbing
M	Mechanical
E	Electrical
T	Telecommunications
R	Resource
X	Other Disciplines
Z	Contractor/Shop Drawings

Table 2.3: Model File Types

Model File Types	
FP	Floor Plan
SP	Site Plan
DP	Demolition Plan
QP	Equipment Plan
XP	Existing Plan
EL	Elevation
SC	Section
DT	Detail
SH	Schedules
3D	Isometric/3D
DG	Diagrams

BIM projects differ from CAD projects in that there may be only one central file representing each discipline’s work, versus the many drawing base and reference files found in the two-dimensional CAD process. However, these model file naming conventions can still be applied. The Model File Type is 3D. On large projects, each discipline’s model may be subdivided for ease of sharing and modification. The user definable portion of the file name can be used to describe the subdivisions, which would typically be by floor or by segment. An example would be “West Wing”:

PROJ-A-3DWEST

As with CAD projects, large BIM projects may require multiple modelers to efficiently complete the design. This process differs between BIM applications but generally consists of a master file with temporary sub-files checked out to individual modelers. For archiving purposes, all sub-files should be saved into the master file, which is then considered the complete BIM file.

Both CAD and BIM files should be saved at key project milestones, such as the end of Schematic Design. They should be maintained in directories that designate the Project Phase, per Figure 2.3. Note that BIM projects may have non-standard phasing. Firms will need to improvise in these cases to accurately communicate the project milestone with which each version of the BIM model is associated. BIM files should be archived both in their native application format and in the Industry Foundation Class (IFC) format.

[1] “Uniform Drawing System,” *National CAD Standard,* Construction Specifications Institute, 2001.

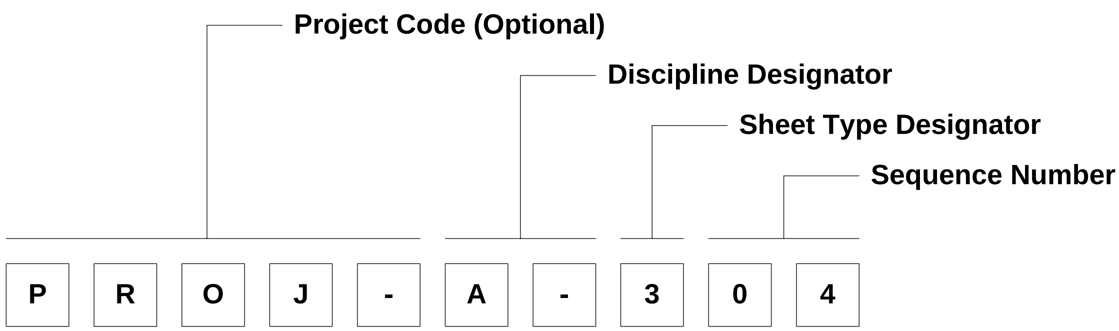
Sheet Files (Outputs)

The sheet file, which contains one or more scaled views of one or more models arranged within a border and title block, is named beginning with the optional five-digit project code followed by:

- Discipline designator
- Sheet type designator
- Sheet sequence number.

See Figure 2.5 for a sample Sheet File and Tables 2.2 and 2.4 for a list of *Discipline Designators* and *Sheet Type Designators*, as published in the *National CAD Standard, Version 2.0*. [2]

Figure 2.5: Sheet File Naming, *Uniform Drawing System*



Source: Kristine Fallon Associates, Inc.

The sheet files would become the outputs in archival PDF format. The naming would remain the same with the PDF file extension.

Table 2.4: Sheet Type Designators

Sheet Type Designators	
0	General (symbols, legend, notes, etc.)
1	Plans (horizontal views)
2	Elevations (vertical views)
3	Sections (sectional views)
4	Large scale (plans, elevations or sections that are not details)
5	Details
6	Schedules and diagrams
7	User defined
8	User defined
9	3D views (isometrics, perspectives, photographs)

Drawing outputs from a BIM are similar to CAD outputs and should follow the same naming conventions.

Summary of Design Firm Submission Recommendations

In summary, the design office should prepare digital design data in the following ways:

- Maintain checkpoint data that corresponds to each major project milestone, documenting the relationship between output and native data at that point
- Follow standard or well-documented digital data organization and naming conventions
- Provide output images, such as renderings and Photoshop montages, in uncompressed TIFF format
- Provide digital drawings, PowerPoint presentations and hybrid outputs in PDF format

[2] "Uniform Drawing System," 2001.

- Provide animations in MPEG-2 format or as individual still frames in TIFF, PNG or BMP format
- Provide interactive 3D content in IFC format or alternatively in X3D or U3D format
- Embed source color profiles and rendering intents in TIFF and PDF files
- Embed all components of compound files—particularly externally referenced files in CAD—in a single file when possible
- Document all linked or referenced files if embedding of components is impossible
- Provide native data in original format.

Digitizing the Existing Collection

Once an institution begins a digital collection, it may become desirable to create digital versions of paper-based documents from the collection as well. This chapter documents best practices for digitizing based on recommendations from the Library of Congress, the National Archives and Records Administration, the Digital Library Federation, Cornell University Library and the *NINCH (National Initiative for a Networked Cultural Heritage) Guide.*

The best practices that follow in this section should be taken as guidelines and should be tailored to the intended uses of the digital images—print or electronic display—and whether enlargement is desired. Below is a list of potential uses and formats.

For print, output formats may include:

- Large-scale printed exhibition images
- Images for publication in books, brochures or papers
- Reproductions of scaled drawings for architects and building owners.

For electronic display, output formats may include:

- Full-screen images for projection in exhibitions or in presentations
- Medium-scale and thumbnail images for Web presentation
- High-quality purchasable online images.

Equipment

The quality of digital capture achieved is directly related to the quality of capture equipment used by the archivist. A digital capture device takes a sample of the analog source material and creates a digital surrogate. Digital capture devices exist for capturing images, text, audio, video or 3D objects. The components involved in digitizing an image include the following hardware and software components:

Hardware:

- Computer, monitor and large data storage device
- Scanner and/or digital camera with copy stand
- Color profiling hardware.

Software:

- Image manipulation program, such as Adobe Photoshop
- Color management software.

Scanners

A multitude of information and research exists on scanners and their various applications. The most important attributes to consider when selecting a scanner are: optical dpi, material handling, size of original accommodated and cost.

The first scanner attribute is the optical dpi, or dots per inch. The optical dpi determines the available range of resolutions for scanned images and therefore the amount of flexibility the user has to enlarge images or save them at a high resolution needed for print. It is important to compare *optical* dpi because scanners will often advertise a higher dpi that is achieved with interpolation. Interpolation is a mathematical procedure that calculates and fills in the unknown values or dots in an image based on the surrounding values or dots. Therefore, the optical dpi is the true dpi.

It is important to match the handling of the documents by the scanner with the type and condition of the documents. Unmounted, flexible architectural plans and renderings in good condition can be accommodated by scanners that require the document to be pulled through the scanning device. Mounted documents, 3D design objects or drawings that are in fragile condition must be laid flat to scan or be photographed digitally.

The scanner must accommodate the size of the expected documents, whether small-scale renderings at 11”x17” or large-format line drawings at 36”x48”.

Some of the most expensive and highest quality scanners produce images that exceed the requirements of the Department of Architecture. Therefore, a balance must be achieved between the quality requirements of the archived images and the cost of equipment.

The two types of scanners that are relevant to the Department of Architecture are sheet fed and flatbed.

Sheet fed scanners, as the name indicates, feed documents through the narrow gap of the scanning device and therefore limit document thickness. Accommodated document thicknesses range from 0.06” to 0.60” for sheet fed scanners. The typical range for optical dpi for sheet fed scanners is 300 to 600 dpi. Monochrome, or black and white, scanners are appropriate for line drawings, while color renderings require a 24-bit color scanner.

For architectural drawings or renderings that are in good condition and of robust materials, a sheet fed scanner can be employed. Carrier sheets should be employed to guard a document during the scanning process. To accommodate large-format architectural drawings, wide format sheet fed scanners are available.

High-end flatbed scanners allow the document to be laid flat and permits edge-matching multiple scans of a document that exceeds the size of its bed. They can be used as an alternative to sheet fed scanners for fine art or documents that are not flexible, too delicate or exceed the size limitation of sheet fed scanners. The Colortrac FB24120 is an example of a high-end flatbed scanner that has an optical dpi of 600, a bed width of 24” and a maximum document thickness of 1”. The Art Institute’s Department of Imaging uses a ScanMate F10 scanner with an optical dpi of 5400 and a 12”x17” bed which would accommodate small-scale renderings but not large-format architectural drawings.

Digital Cameras

As an alternative to scanning, digital cameras provide an option for digitizing works of art, particularly 3D objects. For flat documents, a copy stand setup—with a base to support the document, a column and camera attachment on the column—should be employed. For large-format drawings, it is important to have a copy stand large enough to accommodate drawing dimensions so it can be captured in one image. Currently, The Art Institute’s Department of Imaging uses Phase One PowerPhase FX scanning back on a 30x40" copy stand and must take multiple shots and stitch them together. The Department of Imaging places a sheet of acetate over the drawings to eliminate creases or a sheet of mylar over tracing paper sketches to prevent them from folding.

Outside of a copy stand setup, the Department of Imaging uses a Phase One H20 for 3D objects, a Sinarback 54H for paintings, a Nikon D1X for publicity and a Canon EOS 1DS for location objects and exhibition installations.

Moving images of a 3D object can be created by stringing together a series of still images taken with a digital camera moving around the object or with a fixed camera and a turn table.

3D Digitization

There are methods for creating a 3D digital model from a physical one. A robotic arm with a sensor at the tip traces the geometry of the 3D physical object and builds a digital surrogate. Frank Gehry often uses this technology to create 3D CAD models from physical models. These CAD models can be exported in neutral formats such as IGES and could be archived and viewed using 3D viewers.

Scanning Properties

There are two important characteristics of the image that is taken by a scanner: the sample rate and the sample depth. The sample rate is the scan resolution—the optical dpi discussed above. The sample depth is the amount of information recorded at each sample point. For example, a sample depth of 24-bits captures 8 bits for each of the three color channels (red, green and blue) at each sample point.

Resolution

There are three ways of describing resolution that are often confused with one another:

- **ppi** (*pixels per inch*) refers to on-screen or digital resolution and applies to those *creating* digital image files. The most common screen resolution is 72 ppi, although new monitor technology has produced a screen resolution of 96 ppi.
- **dpi** (*dots per inch*) should be used when talking about printing and refers to the printing dot. Scanners typically use dpi to indicate scan resolution. Many color ink jet printers have a resolvable resolution of 300 dpi. To optimally reproduce an image at a one-to-one ratio, the resolution of the scan should be 300 dpi. High end printers used by magazine publishers will print 600 dpi for glossy documents.
- **lpi** (*lines per inch*) relates to offset and gravure printing and describes the "lines" of the halftone screen. For example, many museum publications are printed with halftone screens of up to 200 lpi. To optimally reproduce an image at 200 lpi, the digital file should have a ppi resolution of 1.5 to 2 times the screen frequency (i.e., 300–400 ppi).

To determine scan resolution, you must know the desired output format of the image. For images to be stored as digital files for on-screen viewing only, 100 dpi resolution is sufficient and will cut down on the file size. For images intended for high-quality print outputs for exhibition or publication, 600 dpi is recommended for black and white line drawings or hand sketches with line strokes where the eye notices sharp transitions from white to black and 300 to 400 dpi for color renderings or images. If conversion to a CAD format is desired by vectorizing the image, a 300 to 400 dpi scan is recommended.

To enable printing at a larger size than the original, the archivist must scan at higher resolution. For example, if the archivist or curator wants to exhibit an 8"x10" rendering at 16"x20" size with a final resolution of 300 dpi, the image must be scanned at 600 dpi. If an 8"x10" rendering is scanning at 300 dpi and then output at 16"x20" size on a 300 dpi printer, interpolation will be performed. With interpolation, a noticeable loss in clarity, sharpness and color occurs. Therefore, the archivist should not rely on this process to make an enlargement, but should use foresight and scan at a higher dpi.

The National Archives and Records Administration suggests less conservative standards for scan resolution. For text, small scale documents are scanned at 300 dpi to work with Optical Character Recognition (OCR) software, used for converting scanned text images to full-text versions. Larger scale text documents are scanned at 200 dpi to save storage space. For images, a standard of 3,000 pixels across the long dimension was set. For maps, plans and oversized records, 300 dpi scanning is used for 11"x17" documents or smaller and 200 dpi for documents larger than 11"x17". With the decreasing cost of storage space, there may be less need to sacrifice resolution for the sake of reducing file size. Table 2.6 explores the relationship between image pixel and inch dimensions.

Table 2.6: Translation of Pixels to Inches with Varying Resolutions

Image Dimensions in Pixels	Image Dimensions in Inches			
	72 dpi	*200 dpi*	*400 dpi*	*600 dpi*
400 x 300	5.6 x 4.2	2.0 x 1.5	1.0 x 0.8	0.7 x 0.5
640 x 480	8.9 x 6.7	3.2 x 2.4	1.6 x 1.2	1.1 x 0.8
1024 x 768	14 x 11	5.1 x 3.8	2.6 x 1.9	1.7 x 1.3
1600 x 1200	22 x 17	8.0 x 6.0	4.0 x 3.0	2.7 x 2.0
3000 x 2250	42 x 31	15 x 11	7.5 x 5.6	5.0 x 3.8
4400 x 3300	61 x 46	22 x 16	11 x 8.3	7.3 x 5.5
6800 x 4400	94 x 61	34 x 22	17 x 11	11 x 7.0
10,200 x 6600	142 x 92	51 x 33	26 x 17	17 x 11
19,200 x 14,400	267 x 200	96 x 72	48 x 36	32 x 24

Source: Kristine Fallon Associates, Inc.

Sample Depth

In addition to the sample rate, choices must be made about the sample depth. This affects the number of bits sampled for each pixel and determines the range of tones captured in the image. Scanners record tonal values as black and white, grayscale and color. In black-and-white capture, each pixel is represented as black or white, on or off. The *threshold* for black can be set. Above this threshold a tone is considered black and below it a tone is considered white. In 8-bit grayscale capture, there are 254 shades of gray along with the black and white. Thresholds can also be set with 8-bit grayscale. In 24-bit color scanning, the tonal values are reproduced with 8 bits in each of three channels—red, green and blue (RGB)—with up to 16.7 million colors. It is important to keep in mind that file sizes for color images are about three times larger than those in grayscale. Some high-end scanners produce images with 48-bit color (16 bits x 3 channels = 48-bit color), such as the ScanMate F10 used by The Art Institute's Department of Imaging. However, this bit depth exceeds the requirements of architectural drawings and bed dimensions of scanners that support 48-bit color tend to be smaller than needed for architectural drawings. Color technology is advancing toward 16-bit depth, confirmed by the expanded support for 16-bit images in the latest release of Adobe Photoshop.

Color technologies have been advancing toward a more complete computer representation of the visible gamut. High Dynamic Range Images (HDRI) present a wider gamut and contrast by adding a fourth color channel to the traditional three (RGB) rather than increasing the bit depth of the existing channels. For example, a 32-bit color space known as RGBE (Red-Green-Blue-Exponent) adds an extra eight bits to the traditional 24-bit color by adding a channel known as "exponent." The role of this fourth channel is to fill in color where the other RGB channels lack. The TIFF format has begun to accommodate RGBE and another HDRI format LogLuv, thereby creating the possibility that current 24-bit images will need mapping to higher ranges in the future. Higher definition computer monitors are required for viewing HDRI.

Because the sample depth has such a drastic affect on file size, it is important to choose what is appropriate to the type of image. The following are tips for choosing the type of sample depth:

- Black and white for line drawings images without shading
- 8-bit grayscale for images with shades of gray or continuous tones such as shaded hand sketches, black-and-white photographs, half-tone illustrations and black-and-white materials where ink density is important
- 24-bit or 48-bit color for images where color is present.

Table 2.7 is a comparison of digitization specifications of various government institutions and research organizations, taken from Cornell University Library findings.

Table 2.7: Comparison of Institutional Requirements for Digitization [1]

Institution/ Organization	Printed Text	Pictorial Materials	Oversized Materials
Library of Congress	300 dpi, 1-bit, TIFF ITU-T.6	3,000 to 5,000 pixels 8-bit gray or 24-bit color, TIFF, uncompressed	[Maps] Color: 300 dpi, 24-bit, TIFF, uncompressed
National Archives and Records Administration	300 dpi, 8-bit gray, TIFF, uncompressed	3000 pixels—long side, 2700 for square, 8-bit gray/24-bit color, TIFF, uncompressed	200 dpi, 8-bit gray or 24-bit color, TIFF, uncompressed
Columbia	600 dpi, 1-bit, TIFF ITU-T.6	200 to 300 dpi, 8-bit gray or 24-bit color, TIFF	[Large format transparency] 4096 x 6144, 24-bit, PhotoCD or TIFF
JIDI (JISC Image Digitization Initiative)	300 dpi, 8-bit (24-bit for color, tinted or discolored originals), TIFF v.6, uncompressed	[Photographic prints] Same as printed text. [Art works] 600 dpi, 8-bit gray /24-bit color, TIFF, uncompressed.	Scan from photo intermediates at 2400 dpi minimum
Memory of the World	200 dpi, 1-bit, TIFF v.6, ITU-T.6	100 dpi, 8-bit gray or 24-bit color, TIFF-JPEG lossless or lossy for non-critical images	100 dpi, 8-bit or 24-bit, TIFF-JPEG lossless. For maps larger than A3, use photo intermediates.
Colorado Digitization Project	600 dpi, 1-bit, TIFF, uncompressed 300 dpi, 8-bit gray, 24-bit color, TIFF, uncompressed	[Photographs] 3000 to 5000 pixels, 8-bit gray/24-bit color or greater, TIFF, uncompressed. [Graphic Materials] 3000 pixels or 300 dpi, 8-bit gray/24-bit color or greater, TIFF, uncompressed	[Maps] 300 dpi, 8-bit gray/24-bit color, TIFF, uncompressed.
California Digital Library	600 dpi, 8-bit gray, TIFF-LZW	600 dpi, 24-bit color, TIFF-LZW	600 dpi if possible, but no less than 300 dpi, 24-bit color, TIFF-LZW

Formats

Once an image has been scanned, choices must be made about file format and file size for storage and steps must be taken to ensure color is reproduced correctly. The recommended format for storing preservation quality digital masters is uncompressed TIFF (Tagged Image File Format). The Art Institute's Department of Imaging uses this format for digital masters and the Digital Library Federation (DLF) confirms the archival application of TIFF in a discussion of *File Formats for Digital Masters* [2]. Uncompressed TIFF retains all the information encoded at the time of scanning, and this is known as a "lossless" image format.

For black and white line drawings where there are large white spaces and patterns of black bits followed by white bits, the use of a lossless compression algorithm is suggested. A lossless compression will store patterns of bit information rather than the individual bits themselves and can therefore greatly reduce file size. TIFF ITU-

[1] Cornell University Library, Moving Theory to Practice: Digital Imaging Tutorial, 2001, available from ***http://www.library.cornell.edu/preservation/tutorial/conversion/table3-1.html***; Internet; accessed 17 September 2003.

[2] Linda Serenson Colet, Don Williams, Donald D'Amato and Franziska Frey, Guides to Quality in Visual Resource Imaging, Digital Library Federation, Council on Library and Information Resources, 2000, publication online, available from ***http://www.rlg.org/visguides***; Internet; accessed 29 January 2004.

T.6 is an example of a lossless compression TIFF format used by the Library of Congress and Columbia for black and white text, a similar application to black and white line drawings. LZW (Lemple-Zif-Welch) is another lossless compression algorithm.

From the uncompressed TIFF or lossless compressed TIFF, derivative files can be saved in formats such as JPG, GIF, MrSid, PNG, and PDF, using an application such as Photoshop. Some formats, such as JPG, use "lossy" compression algorithms that offer a greater amount of compression but must sacrifice data to minimize file size. Examples of lossy compression, in which data are *lost*, are fractal and wavelet compression. The archivist can also reduce the image dimensions or lower the resolution for specific output purposes. For example, an archivist might create thumbnail images used for database searching or a medium-sized image intended for onscreen viewing. The Colorado Digitization Project has created a quick-reference chart that compares specifications for master, access and thumbnail images. See Table 2.8.

Table 2.8: Master and Derivative Image Resolution, Dimensions, Bit Depth, File Type and Compression [3]

	Master	Access	Thumbnail
Spatial Resolution	3000-5000 pixels on the longest dimension	150 ppi	72 ppi
Spatial Dimensions	100% of original	600 pixels on longest dimension	100-200 pixels on longest dimension
Bit Depth	1 bit bi-tonal 8 bit grayscale 24 bit color	1 bit bi-tonal 8 bit grayscale 24 bit color	1 bit bi-tonal 8 bit grayscale 8 bit indexed color 24 bit color
File Type	TIFF	JPEG	JPEG
Compression	none or lossless	JPEG Medium Quality Compression	JPEG, Low Quality Compression

MrSid (Multi-Resolution Seamless Image Database), developed by LizardTech, Inc. of Seattle, uses wavelet-based image compression, which is especially well-suited for the distribution of very large images. The Library of Congress uses MrSid to deliver maps from its collections. In addition to its impressive compression capabilities, it stores multiple resolutions of images in a single file and allows users to select the resolution (in pixels). The National Aeronautics and Space Administration (NASA) uses MrSid as a viewing technique for the collection of satellite images taken by the Landsat Satellite, used to study the earth's environment, resources and natural and man-made changes.

Table 2.9 summarizes the most common raster image formats and their characteristics, according to the *NINCH (National Initiative for a Networked Cultural Heritage) Guide*.

[3] Western States Digital Standards Group, Western States Digital Imaging Best Practices – Quick Reference, January 2003, available from ***http://www.cdpheritage.org/resource/scanning/WSDIBP/quickref.html***; Internet; accessed 2 June 2004.

Table 2.9: Comparison of Common Raster Image Formats [4]

Extension	Meaning	Description	Strengths/weaknesses
.tiff, .tif	TIFF (Tagged Image File Format)	Uncompressed file. Originally developed for desktop publishing. 1 to 64 bit depth. Used mostly for high quality imaging and archival storage.	Generally non-compressed, high quality. Large file sizes. Most TIFF readers only read a maximum of 24-bit color. Delivery over web is hampered by file sizes. Although LZW compression can reduce these file sizes by 33% it should not be used for archival masters.
.gif	GIF (Graphics Interchange Format)	This 8-bit file format has support for LZW compression, interlacing and transparency.	Lossless compression. Popular delivery format on web. .png was defined to replace GIF.
.jpg, .jpeg	JPEG (Joint Photographic Experts Group)	Compressed images. 8-24 bit. Variable amount of compression to vary quality and file size.	Lossy compression. Widely used delivery format. Flexible.
MrSid	Multiresolution Seamless Image Database	Image-compression technology	Lossy compression. Can compress pictures at higher ratios than JPEG; stores multiple resolutions of images in a single file and allows the viewer to select the resolution.
.pcd	ImagePac, PhotoCD	Lossy compression. 24 bit depth. Has 5 layered image resolutions.	Used mainly for delivery of high quality images on CD.
.png	PNG (Portable Network Graphics)	Lossless compression. 24 bit. Replaced GIF due to copyright issues on the LZW compression. Supports interlacing, transparency, gamma.	Some programs cannot read it.
.pdf	PDF (Portable Document Format)	4-64 bit depth. Uncompressed. Used mainly to image documents for delivery.	Need plug-in or adobe application to view.
.pct	PICT	Compressed. Mac standard. Up to 32 bit. (CMYK not used at 32 bit.)	Supported by Macs and a highly limited number of PC applications.

File Size

File size and storage space are a concern for a scanning project with a scope as large as that of the Department of Architecture. The sample depth and the resolution of the scan both contribute to the file size of a digital image. Recall that the sample depth or *bit* depth is the product of a number of bits per pixel and the number of channels. For example: 24-bit color bit depth = 8 bits/channel x 3 color channels. To calculate the file size, one can use a formula based on pixel dimensions or inch dimensions, shown below.

[4] *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, Humanities Advanced Technology and Information Institute, University of Glasgow and National Initiative for a Networked Cultural Heritage, February, 2003, Publication online, available from ***http://www.nyu.edu/its/humanities/ninchguide***; Internet; accessed 29 January 2004.

(Pixel dimensions) x (# bits/pixels) x (# of channels) x (# bytes/bit) = file size
Example: 4,300 x 5,300 pixels at 24-bit color
(4,300 x 5,300 pixels) x (8 bits/pixel) x (3 channels) x (1 byte/8 bits) = 68,370,000 bytes = approximately 69 MB

OR

(Height in inches x width in inches x dpi^2 x bit depth) x (1 byte/8 bits) = file size
Example: 11 inches by 17 inches at 400 dpi, 8-bit grayscale
(11 x 17 x 400^2 x 8) / 8 = approximately 30 MB

Large-scale architectural images in color can easily be 300-400 MB large. Thousands of scanned images from past collections will create a need for many gigabytes of digital storage. The Art Institute's Department of Imaging aims for 200 MB files at 16-bit depth, but sometimes produces files up to 540 MB. Table 2.10 summarizes the relationship between pixels, inches and file size.

Table 2.10: Relationship between Pixels, Inches and File Size
Note: DVD storage value used is 4.7 GB each

Dimensions in Pixels	Image Dimensions in Inches				File Size (Grayscale)	No. / DVD	File Size (Color)	No. / DVD
	72 dpi	*200 dpi*	*400 dpi*	*600 dpi*				
400 x 300	5.6 x 4.2	2.0 x 1.5	1.0 x 0.8	0.7 x 0.5	0.120 MB	39,166	0.360 MB	13,055
640 x 480	8.9 x 6.7	3.2 x 2.4	1.6 x 1.2	1.1 x 0.8	0.307 MB	15,299	0.922 MB	5,100
1024 x 768	14 x 11	5.1 x 3.8	2.6 x 1.9	1.7 x 1.3	0.786 MB	5,979	2.36 MB	1,991
1600 x 1200	22 x 17	8.0 x 6.0	4.0 x 3.0	2.7 x 2.0	1.92 MB	2,447	5.76 MB	815
3000 x 2250	42 x 31	15 x 11	7.5 x 5.6	5.0 x 3.8	6.75 MB	696	20.2 MB	232
4400 x 3300	61 x 46	22 x 16	11 x 8.3	7.3 x 5.5	14.5 MB	324	43.6 MB	107
6800 x 4400	94 x 61	34 x 22	17 x 11	11 x 7.0	29.9 MB	157	89.7 MB	52
10,200 x 6600	142 x 92	51 x 33	26 x 17	17 x 11	67.3 MB	69	201 MB	23
19,200 x 14,400	267 x 200	96 x 72	48 x 36	32 x 24	276 MB	16	829 MB	5

Source: Kristine Fallon Associates, Inc.

Color Management

Color management techniques for digitizing follow the recommendations outlined for design firms, with some opportunities for automation. All hardware devices should be calibrated and an ICC color profile for each should be recorded, as recommended for design firms. The color profile of the source device should be embedded in the digital image. The process of embedding the color profile can be automated by some scanners and digital cameras. Therefore, the additional step of manually embedding a profile may be eliminated. Also, there are fewer source devices—a limited number of scanners and/or cameras as opposed to potentially hundreds of designers' computers—and these devices can be frequently calibrated for best accuracy.

The Art Institute's Department of Imaging uses GretagMacBeth SpectroScan spectrophotometer and Profile Maker for creating device profiles. Digitized images are assigned the custom color profile of the capture device. If color correction is needed, images are converted to a large working space, manipulated and saved with that working space embedded.

Facilities and Resources

For objects subject to conservation, the facility in which the scanning process takes place should be properly climatized, with adequate cooling and air-circulation to counteract the heat from the equipment. There should be a minimum of ambient light, making the primary light source that of the scanner bulb or cold lights for a digital camera. Any lights should be UV filtered or give minimal UV readings. Exposure to the high-intensity light of scanning may be problematic for the documents.

Table 2.11, excerpted from the Digital Library Federation's *Guides to Quality in Visual Resource Imaging*, summarizes personnel roles for a large digitization project.

Table 2.11: Resource Requirements for Digital Imaging Project [5]

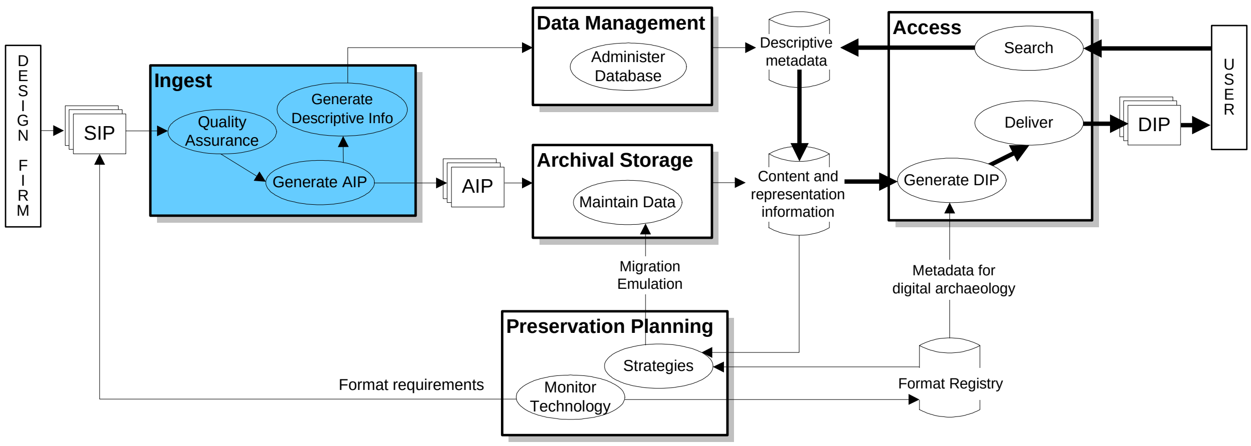
Staffing	Description	In-house project	Outsourced project
Project managers (internal)	Internal project managers are required to manage project goals and institutional expectations, identify staffing and equipment costs, coordinate archival and access needs across departments, and adapt the digital plan as necessary to achieve success.	x	x
Vendor project managers	Vendor project managers run the digital operation and allocate appropriate staffing and expertise to the project.		x
Photo services staff	Institutions with an internal photo services division should use it to manage, operate, and maintain the digital project. If the project is outsourced, photo services staff must closely interact with the vendor.	x	x
Curatorial and archives staff	Internal curatorial or archives staff members, or both, identify project goals, choose objects for digitizing, identify preservation concerns, and project the short- and long-term uses for the digital images.	x	x
Conservators and preservationists	In-house conservators or preservationists should review the project before it begins and identify conservation and preservation concerns.	x	x
External consultants	External consultants may advise on digital studio setup, system integration and networking concerns, archival storage issues, color-management needs, and other matters.	x	x
Grant writers	Grant writers may be needed to write proposals to secure funding for the project.	x	x
Computer/technology staff	Computer and technology staff set up the system, resolve network issues, design storage systems, and performs similar tasks. An institution that does not have these resources must secure outside resources to set up the computers and networks and to handle maintenance.	x	x
Preparators and art handlers	Preparators and art handlers prepare and transport objects to the studio for digitizing. An institution dealing with surrogates may not require this type of staff.	x	x
Quality and production managers	These supervisors set and maintain image quality-control standards and production goals. Their functions should be separate from those of the scanner or camera operators and technicians.	x	
Scanner or camera operators and technicians	Scanner or camera operators and technicians capture and edit the original object or surrogate.	x	

[5] Excerpted from: Colet, Linda Serenson, Don Williams, Donald D'Amato and Franziska Frey, *Guides to Quality in Visual Resource Imaging*, Digital Library Federation, Council on Library and Information Resources, 2000, Publication online, available from ***http://lyra2.rlg.org/visguides/visguide1.html***; Internet, accessed 25 May 2004.

Post-processors	Once a digital image has been captured, it is passed to a post-processor, who processes it on archival storage mediums and prepares it for short- and long-term use.	x	
Administrative assistants	Assistants create and maintain archival logs and keep track of the metadata information to ensure that the digital process is documented and that the documentation can be searched for easy retrieval.	x	x
Vendor services for digital capture, post-processing, and administering logs	Significant vendor costs will be incurred for digital capture, post-processing, administering logs, and equipment use. Often this cost is subsumed in the per-image cost.		x

Collecting and Processing Digital Design Data

Figure 2.1b: Collection and Archiving System: Ingest



Accessioning Process

Most digital repository software packages provide convenient means for individuals to make submissions to the archives and for these submissions to be reviewed electronically and accepted. For example, they allow an individual researcher to upload his or her publications through a Web interface.

In museums, the accessioning of art objects is a lengthy process involving curatorial selection and multiple approvals. For this reason, digital design data will likely be ingested into the system on arrival, but cataloged with a “pending” or “temporary” status. Once approval has been granted, the status of the digital objects can be changed to “accessioned.”

At The Art Institute of Chicago, design drawings enter the collection in two major ways. One or several pieces may be offered to or solicited by the Department of Architecture curators from a specific designer. Occasionally, the Department is offered the entire archive of a design office. In the latter case, the gift usually includes or is accompanied by a grant funding the effort of sorting, evaluating and cataloging the collection. Some pieces are accessioned into the permanent collection; other pieces are archived in the study collection; still others may be discarded.

With a digital collection, the curator will continue to exercise discretion, either by visiting the design office to select digital materials for submission or by reviewing the submission when it arrives. Following the two-tiered collection approach, output data—data that represent the designer’s intent, that are judged to have artistic value and that are in a suitable format for long-term archiving—will be accessioned. The native or source data will not. The accessioned data will be part of the permanent collection and the source data will be part of the study collection.

In the more selective curatorial process, Art Institute Department of Architecture curators will visit the design firm to identify output files of interest and work with architects to identify the associated native data and how files will be organized and named for submission. The designer then will either be given access to a Web site for submission or submit via CD, DVD or other mutually agreeable medium.

Submission and ingestion of digital files should be simple and easy to accomplish. While a public submission utility is not desirable, there are advantages to having a controlled-access Web-based interface through which design firms could upload their files. The institution should determine whether this approach provides sufficiently clear provenance, or whether a technique such as digital signatures should also be required. A digital signature verifies both the identity of the signer of the file and that the contents have not been altered since the signature was affixed.

Because of the nature of digital data, the current registrarial procedure for receiving works into the archive will need to be updated. At almost every institution, the procedure requires receipt of a physical object.

In current practice at The Art Institute of Chicago, a Deed of Gift is completed and signed by the Donor initiating joint and equal copyright ownership between the Donor and The Art Institute for reproduction, creation of derivative works, distribution of copies for sale or other transfer and public display. Digitized versions of the original become the sole property of the creator of the digital surrogate. A Deed of Gift for a digital submission should also include an agreement to allow reproduction of digital design data in any medium known or not yet invented to display, transmit, publish, reasonably adapt or otherwise use.

Finally, ownership of digital data can be unclear. The creator of an electronic design is the owner, unless the design was created by an employee within the scope of his or her employment (work for hire), or the creator has by contract transferred his or her rights in whole or in part to another. With digital designs ambiguity arises because it is impossible to distinguish a copy of a digital file from the original. In some design projects, the client demands ownership. In the old physical world, the original drawings would be transferred to the client, so they would not be available for the design firm to transfer subsequently to an archival institution. With digital design data, the client would receive the data, but an indistinguishable copy would most likely remain on the design firm’s server. In ten years, no one may remember that the firm does not own that design.

Ingest Process for Digital Design Data

The primary goals for Ingest are to check the data for corruption and to transform the Submission Information Package (SIP) to an Archival Information Package (AIP). An Archival Information Package (AIP) is created from one or more SIPs and complies with the archive’s data formatting and documentation standards. This may require a reorganization of the original files and creation of derivative image files. An AIP will contain the following elements:

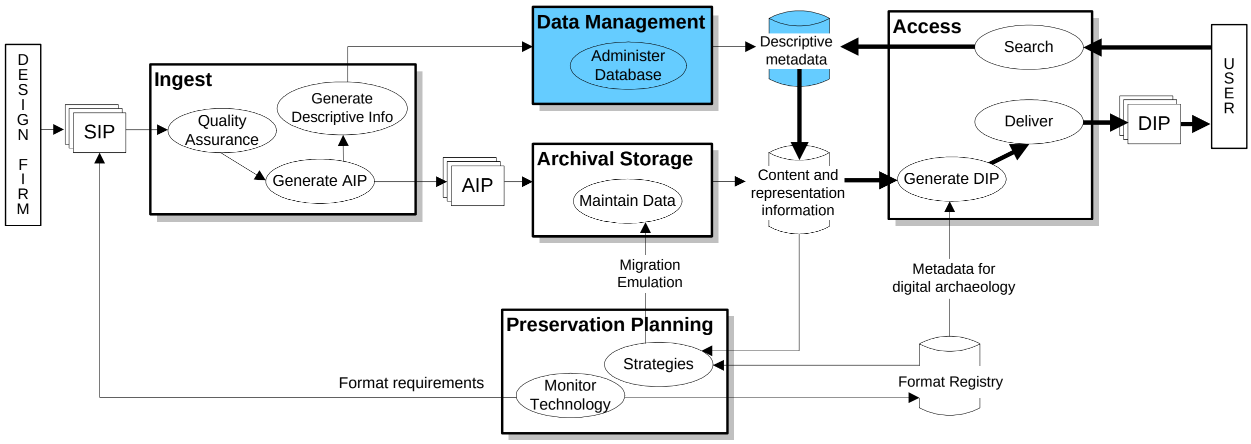
- The digital content in the original format
- Representation information to understand and view the format
- Confirmation of authenticity and data integrity
- Descriptive metadata relevant for searching from the Access module
- Administrative metadata with provenance information and preservation strategies
- Structural metadata, documenting the relationships among the files.

After digital design data have been received by the Department of Architecture, the archivist will perform the following procedures on the data:

- Create an initial catalog record with basic project information (to be sent to the Data Management module)
- Upload digital documents in groups by design phase, preferably with automatic creation of checksum values for all data to ensure their long-term integrity
- Complete quality assurance (QA) checks to ensure no file corruption has taken place
- Assign a “pending” status to all digital documents as they await approval for accessioning
- Create derivative JPG images from the master TIFF images for use on the Web
- Generate an AIP using a programmed routine to bundle descriptive, administrative and structural metadata with the digital content in the format expected by the back-end data repository and send to Archival Storage and Data Management modules.

Cataloging Digital Design Data

Figure 2.1c: Collection and Archiving System: Data Management



The Data Management module maintains information about the digital data—metadata. Metadata are used to organize the information system and to search for particular items in the collection. If the metadata are to be effective and effectively used, it must classify the data in a way that is appropriate to those data. For example, the criteria needed to catalog architectural drawings are different than those needed to catalog a zoological specimen: for both, we want to know where the object came from, but for the architectural drawing we want to know who the creator was, while for the animal specimen, we need to know its genus and species. This chapter will discuss:

- The definition of metadata
- Metadata schema relevant to the Department of Architecture
  - Dublin Core
  - Categories for the Descriptions of Works of Art (CDWA)
- The current Art Institute collection management system called CITI (Collection Image Text and Index) that implements CDWA
- A new Department of Architecture document classification scheme.

Metadata

Metadata are defined as data or information about other data. Metadata are used in library cataloging and have become an integral part of searching on the Internet.

Types

There are three types of metadata as defined by The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials[1]: descriptive, administrative and structural:

- Descriptive metadata identify and describe the information with fields such as creator or artist, title, subject matter and so forth, to facilitate searching, retrieval and management of resources. They include bibliographic information, catalog information and topic information.
- Administrative metadata are used to manage the digital resources and include acquisition and accession information, intellectual property status, preservation information and digitization

[1] *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials,* Humanities Advanced Technology and Information Institute, University of Glasgow and National Initiative for a Networked Cultural Heritage, February 2003, available from ***http://www.nyu.edu/its/humanities/ninchguide***; Internet; accessed 5 March 2004.

specifications such as the hardware used to digitize, resolution, compression and file size (in bytes). Administrative metadata are used track the resources and to aid in preserving them over the long term.

- Structural metadata describe the internal structure of a digital resource and relationships between its components, such as between a PowerPoint presentation and the related image or animation files. They can also relate multiple versions of a resource, such as a high-resolution master image and low-resolution derivative images and thumbnails.

Attributes

For each type of metadata, there is source, status and level information.

- *Source:* The source of the metadata can be internal to the resource—defined at the time the resource was created—or it can be external and be added manually by an archivist. Metadata that are internal to a digital resource include file name, file format and header information with resolution, compression and source color profile for images. Some internal metadata, such as file name, could be automatically extracted to populate a metadata record. Examples of manually entered metadata would be accession information, rights or descriptive information provided by the design firm.
- *Status:* A resource can have metadata with different statuses such as static metadata that never change (title, provenance, date of creation, creation attributes) or dynamic metadata (location, user transaction logs) or long-term metadata used to ensure accessibility of the resource over time (technical format and preservation information).
- *Level:* There may be multiple levels of metadata, for example: collection metadata and individual item metadata. These are especially applicable to the Department of Architecture where there is a hierarchical relationship between a job or project and the individual drawings, images and other artifacts pertaining to that project.

Schema

There are sets of semantics that exist for describing, organizing and searching metadata. Many research institutions and collaborations of librarians, archivists and computer scientists have created metadata schema that define information requirements for cataloging an object or work. Some metadata schema also define a data structure for the metadata. Two relevant metadata initiatives—Dublin Core and Categories for the Descriptions of Works of Art (CDWA)—have different levels of semantic complexity and structural capability.

Dublin Core Metadata Initiative – **http://dublincore.org**

One of the most basic and widely used metadata schema is outlined in the *Dublin Core Metadata Initiative (DCMI)*, developed at a workshop held in Dublin, Ohio in 1995. This workshop was sponsored by the National Center for Supercomputing Applications (NCSA) at the University of Illinois Champaign/Urbana and the Online Computer Library Center (OCLC) to discuss and develop a core set of semantics for Web-based resources that would make search and retrieval of information from the World Wide Web easier. The common set of Dublin Core metadata is designed to allow for discovery of content across digital repositories worldwide through what is called the Open Archives Initiative (OAI).

The Dublin Core standard includes two levels: simple and qualified. The Simple Dublin Core defines an “Element Set” of 15 essential metadata fields for archiving data including author, creator and subject. It was designed to be a “least common denominator” that could be used for basic discovery across as wide a range of digital archives as possible. (See Table 2.12 below for the full list.) The Qualified Dublin Core refines the element set by appending qualifiers to the elements that can be tailored to the needs of the institution, such as “Creator.Architect” or “Creator.Draftsman.”

The Dublin Core developed from a bibliographic point of view and was designed primarily to store and make available written documents (e.g., books, journal articles, research reports and laboratory notes) in a digital form. As such, it has features that are tailored to library users and academic—often scientific—research disciplines. Dublin Core does not accommodate a hierarchical data structure for the metadata, but rather is a flat record.

Dublin Core is mentioned because it is the metadata scheme used by DSpace, the recommended data repository system for the Archival Storage module, as discussed in the *Storing Digital Design Data* chapter. While DSpace could be used to catalog Department of Architecture works, in addition to storing them, its Dublin Core metadata are not sufficient to include the range of descriptive information currently entered for works in the collection. Nor does the Dublin Core accommodate the cataloging hierarchy used by the Department of

Architecture: *Project* → *Drawing Group* → *Individual Drawing*. In DSpace, Dublin Core metadata are linked only at the *Item* level, which is a part of a *Collection*, which is itself a part of a *Community*. Since only the *Item*—the third level in the hierarchy—has associated Dublin Core metadata, it is not possible to search for traits of the Community or Collection. This is awkward for cataloging architectural collections.

Though the Dublin Core metadata scheme is not sufficient to serve as the primary metadata scheme for cataloging a digital design collection, maintaining a secondary copy of the metadata in Dublin Core format is beneficial. The Internet has made it possible to create "virtual" collections that span multiple institutions and DSpace is designed to be a part of this type of federated model of multiple institutions. Having a secondary Dublin Core record for each architectural work would open the collection to a greater audience of online researchers who could potentially conduct searches across multiple institutions. A Simple Dublin Core metadata record also meets the requirements of the Open Archives Initiative (OAI), whose goal is a standard discovery method across data repositories.

Table 2.12 lists and comments on the Simple Dublin Core metadata elements.

Table 2.12: Dublin Core Metadata Elements [2]

Element Name: Title	
Label:	Title
Definition:	A name given to the resource.
Comment:	Typically, Title will be a name by which the resource is formally known.
Element Name: Creator	
Label:	Creator
Definition:	An entity primarily responsible for making the content of the resource.
Comment:	Examples of Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Element Name: Subject	
Label:	Subject and Keywords
Definition:	A topic of the content of the resource.
Comment:	Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Element Name: Description	
Label:	Description
Definition:	An account of the content of the resource.
Comment:	Examples of Description include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
Element Name: Publisher	
Label:	Publisher
Definition:	An entity responsible for making the resource available
Comment:	Examples of Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Element Name: Contributor	
Label:	Contributor
Definition:	An entity responsible for making contributions to the content of the resource.
Comment:	Examples of Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Element Name: Date	
Label:	Date
Definition:	A date of an event in the lifecycle of the resource.

[2] *Dublin Core Metadata Initiative,* Dublin Core Metadata Element Set, Version 1.1: Reference Description, *02 June 2003, available from* http://dublincore.org/documents/dces/dct1#dct1; Internet; accessed 26 May 2004. Copyright © 2003 Dublin Core Metadata Initiative. Status: This is a DCMI Recommendation.

Comment:	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.
Element Name: Type	
Label:	Resource Type
Definition:	The nature or genre of the content of the resource.
Comment:	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.

Element Name: Format	
Label:	Format
Definition:	The physical or digital manifestation of the resource.
Comment:	Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).
Element Name: Identifier	
Label:	Resource Identifier
Definition:	An unambiguous reference to the resource within a given context.
Comment:	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Element Name: Source	
Label:	Source
Definition:	A Reference to a resource from which the present resource is derived.
Comment:	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Element Name: Language	
Label:	Language
Definition:	A language of the intellectual content of the resource.
Comment:	Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian", and "en-GB" for English used in the United Kingdom.
Element Name: Relation	
Label:	Relation
Definition:	A reference to a related resource.
Comment:	Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Element Name: Coverage	
Label:	Coverage
Definition:	The extent or scope of the content of the resource.
Comment:	Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.
Element Name: Rights	
Label:	Rights Management
Definition:	Information about rights held in and over the resource.
Comment:	Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.

Categories for the Description of Works of Art – *www.getty.edu/research/conducting_research/standards/cdwa*

A second metadata scheme called ***C**ategories for the **D**escription of **W**orks of **A**rt* (CDWA) was specifically designed for describing works of art, architecture, groups of objects and visual and textual surrogates. CDWA was developed by the Art Information Task Force (AITF) through the combined efforts of art historians, museum curators and registrars, visual resource professionals, art librarians, information managers and technical specialists. It was funded by the J. Paul Getty Trust with a two-year matching grant from the ***National Endowment for the Humanities (NEH)*** *to the* ***College Art Association (CAA)***. CDWA has become the preferred metadata scheme for use with art and architecture collections.

CDWA defines requirements—27 categories with subcategories—for a metadata scheme for describing and accessing art and architectural objects (see *Appendix B: CITI Implementation of CDWA* for full listing of categories). CDWA does not prescribe a data structure for the metadata, but it does suggest a metadata hierarchy that allows information to be recorded at both a master level and at a component level (or at an architectural job level and individual document level). To aid in creating a Master/Component structure, CDWA provides the OBJECT/WORK – COMPONENTS field that allows an object to act as a master record with many component records. Thus, CDWA accommodates a cataloging hierarchy appropriate to a collection of architectural drawings and other media.

As an example, the institution would collect digital design data from a project. The project would be indexed at the OBJECT/WORK level. For that project, there would be several hundred digital documents such as drawings, renderings and PowerPoint presentations. These would be cataloged as COMPONENTS.

Another useful feature of CDWA is the concept of an "Authority" record that can be linked to many objects or groups of objects to minimize data re-entry. Authorities describe extrinsic information about an art object, namely persons, places or concepts. For example, the Creator Identification authority record is appropriate for describing an architect or architectural firm that can then be linked to all jobs or projects by this architect.

Given that CDWA is the preferred metadata scheme for cataloging architectural works for the Department of Architecture, we must look at the best way to implement it.

CITI (Collection Images Text and Index) Implementation of CDWA

At The Art Institute of Chicago, there exists a collection management system called CITI (Collection Images Text and Index) that implements CDWA as its metadata scheme with a data structure that can accommodate the Department of Architecture cataloging hierarchy. It is therefore the preferred database system to catalog the digital design data collection and manage the database of information. See *Appendix A – CITI Snapshots* to view the current CITI interface. The implementation of CDWA used by CITI may serve as a template for other institutions interested in creating a database system based on CDWA (see *Appendix B – CITI Implementation of CDWA*).

Customization for Department of Architecture

At this time, CITI is not a perfect fit for cataloging Department of Architecture works, but requires customization in three areas: additional metadata tables and fields, refinement of the Master/Component data structure and reduction of redundant data entry.

Additional Architecture-Related Metadata Tables and Fields

The first area of customization in CITI is the addition of tables, fields and terms to accommodate the Department of Architecture cataloging. CITI administrative metadata should include a table recording the history of the translation or migration steps taken to preserve the original digital data (as explained in the *Preserving Digital Design Data* chapter). The information recorded in this table should include:

- The document ID
- The type of preservation action (such as upgrade to a newer software version or conversion from one format to another)
- The state before and after the action (such as previous and new software version numbers)
- The date of the action

- The name of the person carrying out the action.

The following architectural metadata fields should be added to CITI. Some of these fields are already planned additions:

- Building Name
- Building Type
- Building Complex
- Job Number
- Method of Representation or Point of View.

Also, some terms will need to be added to CITI fields. For example, to the Role field, the term "Contractor" should be added and to the Method of Representation field, the following terms should be added:

- Plan
- Section
- Elevation
- 3D View
- Perspective
- Isometric
- Rendering.

Master/Component Data Structure

The second area of customization relates to the Master/Component data structure. The Master/Component data structure should accommodate the following three levels of metadata hierarchy:

- Job/Project level
- Document Group level
- Individual Document level.

The Job/Project level and Document Group level metadata will be entered during Ingest, while metadata for the individual document records will be entered after documents have been approved for accession.

Inheritance Tool

Currently, CITI duplicates metadata fields in Master and Component records. This requires redundant data entry. To eliminate redundancy in metadata entry, Master/Component hierarchies should allow metadata to be "inherited" from one level to the next. CITI programmers have already designed an inheritance "tool" for data entry but it is not scheduled to be built until 2005.

Mapping to Dublin Core Metadata

After cataloging Department of Architecture documents in CITI, selected fields of the CDWA metadata should be automatically mapped to Dublin Core metadata and stored in the back-end DSpace repository. This will enable the Department of Architecture collection to be OAI-compliant and searchable across multiple institutions.

New Document Classification Scheme of the Department of Architecture

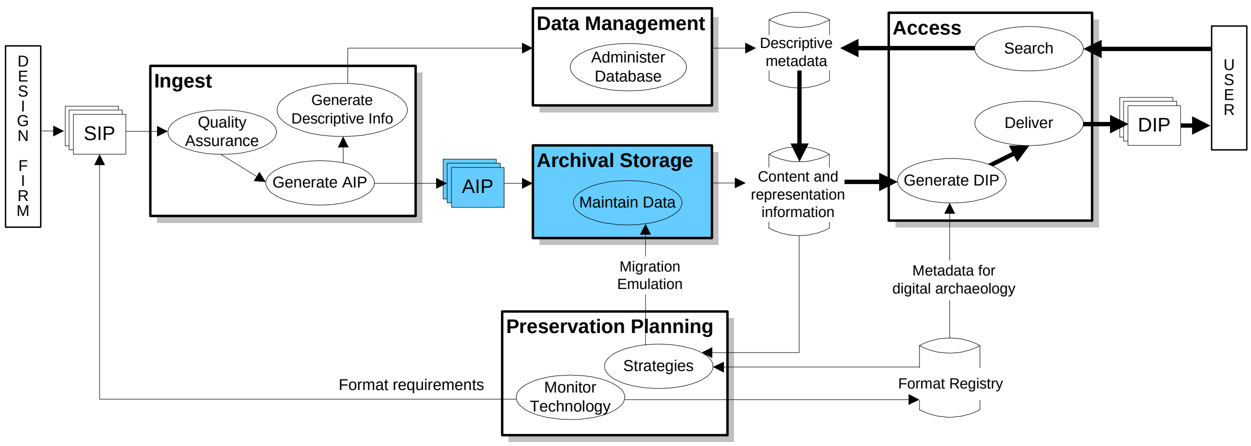
The current classification system used by the Department of Architecture is a based on the type of architecture drawing: presentation drawings, design and detail drawings, working drawings and reference drawings. This classification system is closely tied to the medium of the drawing. Due to the abundance of digital formats, the new classification scheme for the digital collection focuses on project or lifecycle phase, rather than on format. Table 2.13 compares the classification system for the current paper-based collection with the proposed classification for a digital collection.

Table 2.13: Classification Systems for Current and Future Collections

Current Paper-Based Collection	Future Digital Collection
Drawing Type	***Project Lifecycle***
Presentation Drawings Design/Detail Drawings Working Drawings Reference Drawings	Design Programming/Pre-Design Schematic/Conceptual Design Design Development Construction Construction Documents Record Documentation Life Facility Management

Storing Digital Design Data

Figure 2.1d: Collection and Archiving System: Archival Storage and Archival Information Package (AIP)



The Archival Storage Module is the repository that maintains the digital content itself. This back-end repository must do the following:

- "House" the digital collection
- Maintain a persistent unique ID for all digital items as well as the metadata specific to digital objects, such as color profile.
- Ensure bitstream preservation of the digital documents
- Perform functional preservation of the data using preservation strategies determined by the Preservation Policy Committee
- Maintain or link to a format registry to track file formats, versions and associated preservation strategies
- Ensure proper data backup
- Provide for disaster recovery.

There are a number of open-source repository software systems available. "Open-source" means that the software—both the executable program and the original source code—can be freely distributed and modified. Any modifications must also be open-source. This is an appropriate model for educational and cultural institutions because it allows them to build on other institutions' efforts and thereby leverage the combined investment in system development.

Archival Storage Software

Of the currently available open-source systems, DSpace's specification provides the strongest features for maintaining the integrity of the data in an archival sense. One of the stated primary goals of DSpace is to preserve the digital information submitted to it: "DSpace will at a minimum remember the bit sequences in the bitstreams associated with a submission and be able to return those bitstreams to future users. Achieving at least this service level is of critical importance." [1] It also has a modular structure and a well-defined Application Programming Interface (API) that facilitates using parts of the program—such as the data storage and integrity routines—in a customized system. Since The Art Institute will want to use CITI as the search and access mechanism for all collections, the Archival Storage Module will require custom programming to interface with CITI.

[1] Margret Branschofsky et. al., *DSpace Internal Reference Specification*, (Cambridge: Massachusetts Institute of Technology, 2003), Specification online, accessible from ***http://libraries.mit.edu/dspacemit/technology/functionality.pdf***; Internet; accessed 29 January 2004.

DSpace was developed by Massachusetts Institute of Technology (MIT) and Hewlett Packard as a digital repository to capture the intellectual output of multidisciplinary research organizations. DSpace stores digital data—in any file format—as bitstreams along with descriptive and administrative metadata about the digital object, using a Dublin Core scheme. There is a project underway at MIT to implement DSpace as the archival repository for the OpenCourseWare system that contains online course information. This effort will provide a model for how a front-end system such as CITI might communicate with DSpace as the back-end repository.

DSpace provides preservation of the bitstream—the sequence of bits in a file. For each bitstream maintained within the system, DSpace generates and stores an MD5 checksum[2] that can be used to verify the integrity of the stored bitstream over time. DSpace further provides for the long-term physical storage and management of the bitstream in a secure repository and includes standard procedures such as backup, mirroring, refreshing media and disaster recovery. It assigns a persistent unique identifier to each contributed item, and associates this identifier with the item's metadata, to ensure that the item is retrievable. The DSpace storage manager is fully transaction-safe, meaning that should anything go wrong in attempting to add a document, the storage is aborted, ensuring the validity of records in the document database.

DSpace also has a system of functional preservation based on format and tracked by a Format Registry. The Format Registry contains format, version and mimetype information, as well as preservation status (supported, known or unsupported), for each bitstream stored within the system. This is discussed further in the *Preserving Digital Design Data* chapter.

Hardware Considerations

The Archival Storage Module is the home of the digital archive. Whenever a curator is reviewing the digital collection in planning an exhibition, or a scholar is accessing the works of a specific architect, or a high school student is looking for illustrations for her paper on Daniel Burnham, the Archival Storage Module is being accessed. The hardware on which the repository resides and the associated communication components must therefore be sufficient:

- To store the anticipated amount of digital data
- To handle the number of concurrent users anticipated
- To process requests with minimal wait time
- To provide a suitable level of reliability and uptime
- To ensure the security of the archive data.

In designing systems, there is always a trade-off between cost and reliability and performance. Each institution must determine the appropriate level of investment in the IT (information technology) infrastructure for its digital collection, keeping in mind that the data *are* the collection.

There are two aspects of reliability:

- Availability of the data to the end users
- Assurance that, no matter what disaster might befall the institution, the data will be recoverable.

Availability of Archived Data

First, we must consider what might make data *un*available: the server could go down, a hard drive could fail or the connection to the Internet could go down. Until the problem is fixed the digital archives would be inaccessible. To ensure availability, a number of steps can be taken.

- *Server:* If a component of the computer system fails, it might cause the server to become unavailable. If this occurs, the user might need to wait until new hardware components are shipped to the site. The data stored on the server will generally be unaffected by the server failure and can be retrieved once access to the server is restored. To protect against server failure, a mirrored server can be employed. With a mirrored server, a copy of every file saved to one server is automatically saved to the second server. One server fails, the second server will be accessed.

[2] A checksum is a form of digital signature or fingerprint that is calculated from the specific sequence of bytes in a file. Any change to this sequence will result in a different checksum. If a new checksum that is calculated when the file is retrieved is the same as the checksum stored with the file, you can be assured that the file is unchanged.

- *Data:* If a hard drive fails, data itself can be lost, rather than merely *access* to the data. To protect against hard drive failure, RAID (Redundant Array of Independent Disks) technology can be used. RAID technology writes data to multiple drives so that a single disk failure will have no impact on data availability. Mirroring is a type of RAID in which all data on one drive are duplicated in their entirety on another.
- *Internet:* To access data in an online repository continuous connection to the Internet is required. To protect against loss of Internet access a second Internet connection can be installed. In the case that one connection fails, the second will take over. Hardware is also available to aggregate or combine the bandwidth of two different Internet connections. That way, users can have the benefits of a higher speed connection while also protecting against Internet failure.

These measures would provide a high level of availability, but at a cost. The curators, in conjunction with the information systems department, must decide what frequency and duration of downtime are acceptable.

Backup and Disaster Recovery

The Archival Storage Module will contain digital originals. In order to protect this digital data as art originals adequate data backup and disaster recovery provisions must be in place.

The purpose of backup is to create a second copy of the data in case the original copy is erased or corrupted. The backup policy for the digital collection must be established jointly by the curatorial department and the information services department and executed by information services.

Backup policy and procedures should be specifically directed to the protection of the digital design data as original works of art. However, it should be made clear to the donors of the digital files that they are subject to the same deaccessioning policies as other works in institution’s collection. If all or part of a collection is deaccessioned, it will be deleted from the repository. Therefore, the donors should be informed that they cannot depend of the repository to maintain their data for them.

The archival storage system should use RAID storage technology to guard against data loss. Using RAID technology, however, does not entirely eliminate the need for backup because multiple drives can fail simultaneously (for example, due to a severe electrical surge), thus wiping out the data. Backup copies should be maintained at a separate facility from the active servers. Today’s practice is typically to write data to external media, such as magnetic tape, compact disk or DVD. The emerging practice is to write the backup copy to on-line active storage.

Backup copies should be remade periodically based on:

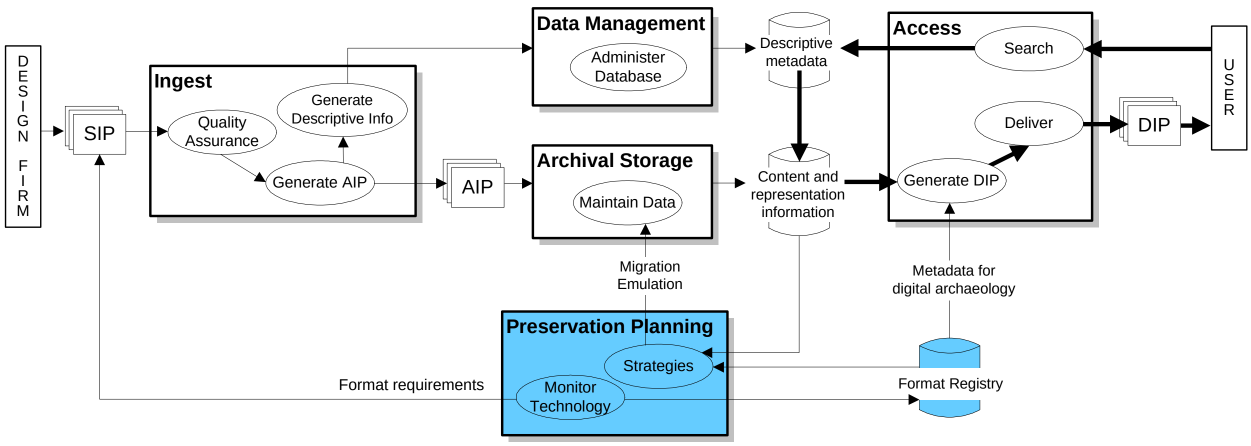
- Any additions to the collection
- Rated life of the backup medium
- Change in selected backup medium, as technology evolves.

Beyond insuring that a current backup copy of the data is maintained, a Disaster Recovery Plan is required for the digital collection. In the case of a natural disaster or act of war, the entire system and the facility in which it is housed could be destroyed. There needs to be a Disaster Recovery Plan that describes how the institution will recover its digital archives and the systems (hardware and software) that make the data accessible. The Art Institute of Chicago duplicates several of its systems at a disaster recovery site. Data are synchronized between the active and disaster recovery systems. However, this option is extremely expensive.

Another option is for one institution to serve as the disaster recovery site for the archives of another institution. For example, The Art Institute of Chicago and the Getty might serve as disaster recovery locations for one another. DSpace is intended to be a federated model that would enable many museums or research institutions to have their collections linked and searchable through one interface and could also facilitate disaster recovery between institutions.

Preserving Digital Design Data

Figure 2.1e: Collection and Archiving System: Preservation Planning



Data preservation is a highly complex issue. Traditionally, paper-based preservation has focused on preserving the physical entity. With digital data, preserving the physical media on which the data are stored solves only part of the problem. Digital preservation requires not only refreshing the physical media and ensuring that it can be read, but also ensuring that the digital data are not changed or corrupted, and maintaining programmatic access to the data.

Preservation Issues

Media refreshing addresses the problem of deterioration of the physical media on which the digital design data are stored. Examples of media refreshing would be copying data from old magnetic tapes to new ones, or replacing the file server's hard drives every 3-5 years and copying all data to the replacement drives.

Ensuring that a data file is not changed or corrupted during storage or transfer can be handled by techniques such as checksums or digital signatures. This aspect is called “bit preservation.”

Because hardware devices, operating systems and application software obsolesce rapidly, the more difficult issues are the availability of hardware that can read the media and of software that can display the content. Maintaining such access to the digital content is known as “functional preservation.”

Archiving the data in active, online storage rather than on external media best solves the media problem. Maintaining access to the variety of native data formats likely to be found in the Department of Architecture's collection poses a greater challenge.

Archival Formats

PDF and TIFF (uncompressed) have been identified as archival formats for output data. The recommended approach is for the designer or firm donating the material to submit output data in these formats. These formats are publicly documented and widely utilized for archival purposes. They are also backwardly compatible, which means that software that can read the current version of the format is also capable of reading all previous versions. This is a major advantage in digital preservation. It will be important to continue to monitor the evolution of these formats going forward. In spite of best intentions, technological change may, at some point, make it impossible to maintain backward compatibility. If this should occur, archival institutions will need to act to preserve their data functionally. However, many institutions will be seeking tools for bringing forward archives, and the market will respond by creating those tools.

Preservation Techniques

DSpace, the recommended repository software system, addresses digital data preservation elegantly. First DSpace identifies two levels of digital preservation:

- ***Bit preservation*** ensures that the file remains exactly the same over time—not a single bit is changed—while the physical media evolve around it. When a file is uploaded to DSpace, a MD5 checksum is generated, reflecting the exact content of data present in the file. The checksum value can be used by downstream preservation services to verify the integrity of the stored bitstreams over time.
- ***Functional preservation*** ensures that the material continues to be usable in the same way it was originally, even though the digital formats and the physical media evolve.

Then DSpace classifies the digital data into three types of formats for preservation purposes:

- ***Supported*** formats are those for which functional preservation can be assured, primarily because the format specification is in the public domain. Supported file formats include PDF, XML, TXT, HTM, JPG, GIF, PNG, TIFF, RTF and Postscript.
- ***Known*** formats are those proprietary or binary formats which are so popular that migration tools are likely to be provided by the software vendors or third parties, thus maintaining functional preservation. AutoCAD DWG is a good example.
- ***Unsupported*** formats are those that are not known and for which functional preservation is not possible. This category is more of an issue in the research community than in design practice.

For all three preservation types, DSpace provides bit-level preservation. The original file should always be preserved so that “digital archaeologists” of the future will have the raw material available for research.

Functional Preservation

Functional preservation of digital data formats requires one of three strategies: migration, translation or emulation. In all cases, the original data file should also be preserved.

Migration

Migration entails conversion of data to new file versions or different formats as the original version or format becomes obsolete. The purpose of migration is to maintain data accessibility over time. Migration requires an ongoing periodic effort to monitor the evolution of the file formats represented in the archive and to convert obsolescing digital objects to current versions and formats. This can be facilitated by using automated tools. Migration usually requires new versions of the proprietary software that created the digital file and does not guarantee a perfect transfer of data: some attributes of the digital object might be lost during the update process.

The “migration on the fly” strategy involves developing conversion tools and programs to translate an obsolete format to a current one, but does not migrate the format immediately. Instead, the institution waits until there is a need to view the obsolete format, at which time it uses the prepared conversion tools to do so. This is a more economical approach than mass migration because only one version of the data is stored, rather multiple files in obsolete and current formats. It begs the question of obsolescence of the migration tools.

Translation

Translation involves moving the data to be preserved to a preferred archival format. In the case of output data, this report recommends the use of PDF and TIFF formats. These formats are capable of capturing all features of the digital output.

Is a similar strategy possible for preservation of the native data? Here the CAD data pose the greatest challenge. There have existed for some time neutral formats for CAD data. A neutral format is a data representation that is not proprietary and is publicly available and documented. These formats are intended to be archival formats and are also used to translate CAD data between proprietary CAD systems. Neutral formats are either official standards or *de facto* industry standards.

So why not translate *native data* into a standard format? This approach raises some distressing questions about the digital “original” and is not recommended. While the output data view is intended to show an explicit set of design elements in a particular way, and this expression can be completely and accurately captured in PDF or TIFF format, the native data serve many—possibly not apparent— purposes and may be the source for multiple and varied outputs and analyses. Further, they may contain non-graphic properties, such as product specification or cost information that were important elements of the design or the design decision-making. They may incorporate journaling, which records the detailed process of creating and modifying the building description. Translating native data into another format will invariably strip them of some attributes and nuances, which can never be recaptured. Even removing data from the software environment that created them, for purposes of viewing, raises questions. As Advisory Committee member William J. Mitchell has written, "Tools are made to accomplish our purposes, and in this sense they represent desires and intentions. We make our tools and our tools make us: by taking up particular tools we accede to desires and we manifest intentions." [1]

Faced, however, with the imminent obsolescence of a particular software environment or with the need for providing access (viewing) of a data format for which there are no readily available viewers, a reasonable functional preservation strategy might be translation of those data into another format. In these cases, however, the original native data should be preserved at the bit level, for the benefit of future researchers. In addition to preserving the bits, it is critical to document in detail the hardware, operating system and software application version in which the data were created. This is the role played by a format registry, as discussed below.

A second and more promising type of translation is to export the native data into a format that is a non-encrypted, legible expression of the proprietary native format. This usually requires the cooperation and consent of the owner of the proprietary format. For example, Autodesk's DXF format served for many years as a complete text representation of the AutoCAD data format. As long as such an export format contains a complete representation of the native data and the export format is documented, it is a highly preferable translation. However, viewers may not be available for the translated data.

Two alternatives to a text file for this type of data translation are relational databases and XML (eXtensible Markup Language) encoding. Relational databases describe the objects within the model in a series of linked tables with fields for all object properties. XML is relatively new tool that is playing an increasingly important role in the exchange of data on the Web. It is derived from SGML (Standard Generalized Markup Language: ISO 8879) and uses tags to describe objects and their properties. These tags are similar to the familiar HTML tags for Web pages but describe the content, not the format. With XML, the programmer defines the tags and the structural relationships between them. The resulting specification is called an XML schema. An XML schema is used to document and standardize the use of the XML tags for a particular purpose. As an example, in 2001 Autodesk released a schema for an XML representation of the AutoCAD version 2002 data format, called DesignXML. This is equivalent to the familiar text-based DXF format, but encoded with more current and flexible technology.

Neutral Formats

The following is a discussion of some of the neutral formats currently available and their limitations in translating the complete content of native data.

IGES

The Initial Graphics Exchange Specification (IGES) is a neutral exchange format for 2D or 3D computer graphics. The need for a common translation mechanism such as IGES arose at a 1979 conference of CAD vendors who were unable to share data among their various CAD tools. IGES presented the first specification for CAD data exchange, published in 1980 as a NBS (National Bureau of Standards, now National Institute of Standards and Technology) report in the U.S.

The IGES file format describes the model as a file of entities. Each entity is represented in an application-independent format to and from which proprietary CAD systems can map their native data representations. IGES therefore has become a translation format between various CAD systems. For example, Doug Garofalo used IGES to translate the structural ribs of the Manilow House from Maya to MicroStation. IGES has also been used to translate UNIX-based CATIA CAD data to Windows-based Rhinoceros CAD to facilitate four-dimensional modeling for Frank Gehry's Walt Disney Concert Hall in Los Angeles and Ray and Maria Stata Center at MIT.

---

[1] William J. Mitchell, *The Reconfigured Eye* (Cambridge: The MIT Press, 1992), 59.

The IGES model is defined with both geometric and non-geometric information. The geometric information consists of points, curves, surfaces, and solids while the non-geometric information includes dimensions, notation, text and grouping information. However, it does not include lighting, view parameters, color or material attributes.

IGES is an aging format and software vendors can be expected to drop support for it as the better and more complete XML options emerge.

ISO / STEP

The efforts toward IGES specifications, done under the auspices of the National Institutes of Standards and Technology (NIST) and the American National Standards Institute (ANSI), were absorbed into the ISO 10303 Standard for the Exchange of Product Model Data (STEP). STEP is a comprehensive ISO (International Organization for Standardization) standard that describes how to represent and exchange digital product or building model information.

The goal of ISO / STEP is to describe digital design data that can span the entire project lifecycle. This includes geometry, topology, tolerances, relationships, attributes, assemblies and configurations. Because the amount of information possibly encoded in a CAD model is constantly changing as technology evolves, it is impossible to develop and maintain a single neutral format to accommodate it all. ISO / STEP uses a technique called application protocols, which limits the purposes or activities supported by the data. An application protocol defines the information requirements for a particular application, or *use*, of the data model. An example of an application protocol is AP 225 for *Structural Building Elements Using Explicit Shape Information*. The result of each application protocol is a neutral format needed to translate intelligent building models from one CAD system to another for specific uses or activities.

There are three elements to every application protocol:

- Application Activity Model (AAM), which describes the supported activities for the building model
- Application Reference Model (ARM), which describes the pieces of information about the building model needed for the activities
- Application Interpreted Model (AIM), written in the information modeling language EXPRESS, which captures the information in the ARM and ties it to a library of pre-existing definitions.

Application Protocol 225 for *Structural Building Elements Using Explicit Shape Information* addresses the exchange of building information between architecture, engineering, and construction application systems. AP 225 includes:

- Three-dimensional shape of building elements
- Spatial configuration of building elements in an assembled building
- Enclosing and separating elements of a building
- Service elements such as plumbing, duct work or conduits
- Fixtures such as furniture and doorknobs
- Equipment such as compressors, furnaces or water heaters
- Spaces including rooms, access areas and hallways
- Specification of properties of building elements, including material composition
- Classification information such as cost analysis, acoustics or safety
- Changes to building element shape, property and spatial configuration information.

AP 225 does NOT include:

- 2D shape representation or drafting presentation
- Implicit representation of building elements by standard parameters
- Structural or thermal analysis
- The assembly process, joining methods or detailed connectivity of building elements
- Building maintenance history, requirements or instructions
- Revisions or design change history
- A bill of materials.

To clarify Application Protocol 225 further, let us examine the example of an element of an intelligent building model: a door. Over its lifecycle, many different parties, including the architect, the permit reviewer, the cost estimator, the procurement group, the installer, and the facility manager will need information about the door. The neutral format created by AP 225 would accommodate the needs of the following parties:

- The architect needs to know the spatial configuration to understand traffic patterns in order to place the door properly: AP 225 encodes information about spatial configuration and about spaces (rooms, hallways, and so forth).
- The permit reviewer needs to know the door's fire rating: AP 225 supports such properties.
- The installer needs to know the hardware set: AP 225 accommodates fixtures.

The AP 225 model would NOT accommodate the following needs:

- The architect's need to know the occupancy type
- The cost estimator's need for a complete bill of materials
- The procurement group's need to know the manufacturer
- The facility manager's need to know the recommended cleaning products for the door's finish.

No application protocol is sufficiently comprehensive for the archiving purposes addressed in this report.

Recent ISO / STEP efforts focus more on a concept called "templates." The contents of a reference data library are determined by the Application Reference Model. There is an unambiguous definition and a specific set of properties for each item in the library. However, organizations may develop "templates" for a specific data set that draw from multiple reference data libraries. This would allow an institution to augment the AP 225 library items with those drawn from other standard or custom-developed reference data libraries.

This approach is powerful, but also complex and immature. As of this writing, the use of templates is in the test bed stage. However, the template approach may provide a very attractive option for functional preservation in five to ten years, as commercial implementations become available.

International Alliance for Interoperability (IAI)

The International Alliance for Interoperability is an alliance of organizations in AEC and other industries whose goal is to develop a universal standard for information sharing and interoperability of intelligent digital building models developed in object-based systems throughout all phases of the building lifecycle. The IAI concept is similar to that of ISO / STEP and IAI is now coordinating its efforts with the ISO / STEP activities.

The IAI has drafted a series of Industry Foundation Classes (IFCs), with specifications that define an object-based data model for the AEC industry. Similar to AP 225, discussed above, the IFC 2x includes the following units of functionality:

- Geometry (volume, areas)
- Building elements (walls, openings, stairs, doors)
- Spaces and spatial structure (space, building story, building site)
- Equipment (ducting, piping, fans)
- Furniture (furniture items, furniture systems)
- Costing (cost planning, estimates, budget).

IFC 2x adds the following functionality, beyond the scope of AP 225:

- Topology (element connectivity, schematic design)
- Relationships between building elements (wall connections, zones)
- Work plans (schedules, resource allocation)
- Orders (work orders, change orders, purchase orders)
- Asset identification (maintenance history, inventories)
- Associated documents
- Actors (people, organizations, addresses).

One difference between the IFC specifications and those of ISO / STEP is that the IFC includes greater entity definition for visualization such as surface style renderings and materials and lighting specifications. For example, surface style rendering is defined by: transparency, color, reflectance, displacement (texture map) and coverage components. An IFC Version 2.0 viewer is available.

Building Lifecycle Interoperable Software (BLIS) is a project to implement IFC standards through a set of use cases, analogous to the application protocols for ISO / STEP. BLIS currently coordinates 60 vendors who seek to support IFC specifications. The DESTINI software under development by BECK, one of the case studies from *Section 1: Current State of Digital Design Tools and Data*, is compliant with the BLIS views of IFC version 2.0.

XML Standards Initiatives - *http://www.w3.org/XML* and *www.xml.com*

As discussed above, XML is emerging as a very popular and flexible tool for data exchange. It permits a programmer to create a specification, or scheme, that defines the tags and the structural relationships between them. In the previous context, XML was suggested as a way of legibly describing proprietary data formats.

However, there are efforts underway to create standard XML schema for the AEC industry. The International Alliance for Interoperability (IAI) has an ifcXML initiative to create XML schema that correspond to the Industry Foundation Classes (IFCs). ifcXML version 1.0 was released in mid-2001. govXML is a proposed subset of the ifcXML standard focused on interoperability in plan review, permitting, inspection and GIS. IAI has also adopted the aecXML initiative, inaugurated by Bentley Systems in August 1999. aecXML shares limited common building components and commercial information between disparate software packages used in the building industry for specific commercial transactions, such as proposals, estimating and scheduling. It is likely that commercial implementations of the ISO STEP template concept will use XML.

Because multiple schema, or namespaces, can be used in a single XML file, standard schema could be augmented by additional namespaces to create a very complete preservation format.

Emulation of the Environment

Emulation means building hardware/software configurations to recreate functionalities of obsolete hardware and software and allow the content of digital data to be used in its original format[2] with the original software. One advantage of emulation is the ability to use the original data preserved as a bitstream. The original "look and feel" of the software application is also maintained. Although it is technically difficult and expensive, emulation is an attractive preservation approach for digital data. However, it requires highly skilled computer programmers to write emulator code, as well as attention to intellectual property and copyright issues related to using or emulating proprietary hardware functions and operating systems. It also poses the challenge of navigating the obsolete user interfaces of original software. For example, a researcher in 2050 would need to learn commands for an MS-DOS system to access records from the mid 1980s or to recognize the "mouse clicks on icons" for a Windows system to access records from the late 1990s.[3]

It is hoped that a governmental body or international consortium will develop emulation environments and make them available to researchers interested in particular sets of data in obsolete formats. There are emulation research and experimentation efforts underway and there are successful emulators for some hardware and operating system environments, such as the Digital Equipment Corporation PDP-series computers and the CP/M operating system.

Format Registry

The Format Registry module in the OAIS reference model is designed to aid in data preservation and to monitor formats. The Format Registry identifies all file formats stored in the archive and their properties and assigns preservation strategies. For example, the Format Registry implemented in DSpace at MIT defines three levels of preservation: supported, known and unsupported.

Besides assisting in the preservation of the digital data, the Format Registry is the source of information for determining the access mechanism for a particular set of data. For example, it would associate the "PDF" file type with the free Adobe Reader.

---

[2] Jeff Rothenberg, *Avoiding Technological Quicksand,* Council on Library and Information Resources, January 1999, Book online, accessible from ***http://www.clir.org/pubs/reports/rothenberg/ contents.html***; Internet; accessed 29 January 2004.

[3] Heslop, Helen, Simon Davis and Andrew Wilson. *An Approach to the Preservation of Digital Records.* National Archives of Australia. December 2002, available from ***http://www.naa.gov.au/recordkeeping/er/digital_preservation/green_paper.pdf***; accessed 19 January 30, 2004.

Global Digital Format Registry

There is an initiative at Harvard University and MIT, with funding from the Digital Library Federation and participation from the Library of Congress and the National Archives and Records Administration, to create a Global Digital Format Registry (GDFR)[4]. The Global Digital Format Registry is a project to create a single, universal format registry to serve multiple repository systems. The GDFR, still in preliminary design, will include:

- Format identification
- Characterization of properties associated with a format
- Rendering or viewing of a format
- Obsolescence assessment risks of a format
- Translation methods from one format to another.

The GDFR has developed an extensive and comprehensive listing of information to be maintained about each data format. This is detailed in the *Appendix E: Global Digital Format Registry*. The hope is the GDFR will eventually serve as a universal registry, linked to most repository software systems and shared by multiple archival institutions.

Active Preservation Management

In the recommended archiving strategy, all original files received from the design firm would always be preserved at the bit level.

The output data in open formats such as PDF and TIFF would require minimal functional preservation because of the backwards compatibility of standard formats. To eliminate access glitches it would be advisable to migrate these standard formats to the most current version of the specification by loading the file into the authoring software and saving it in the latest version of the format. This could be done periodically, rather than each time a new version of the format is released. The process should be automated to eliminate human error, with attention to compression and color management settings.

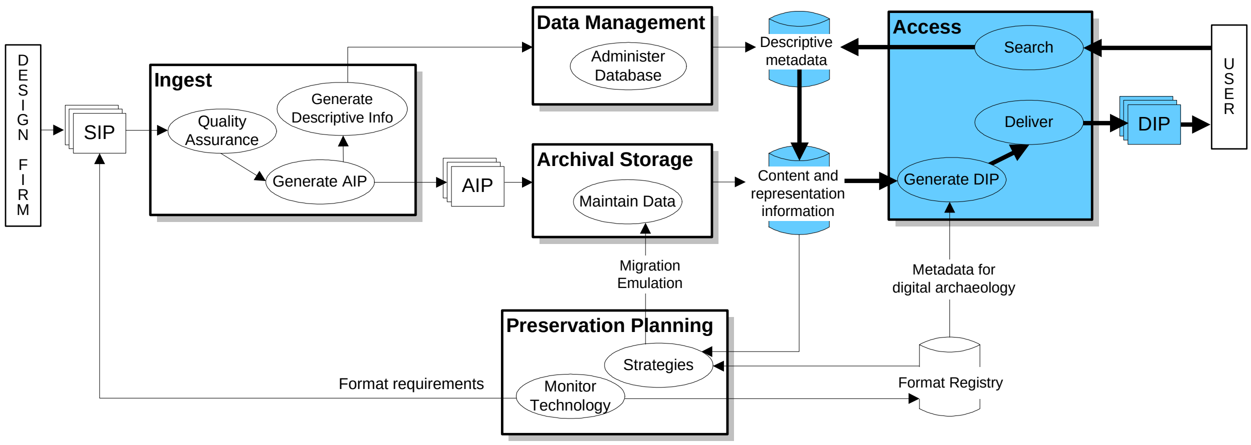
There is an additional functional preservation consideration for PDF files containing embedded animations. Embedding these animations only embeds the *content* in the PDF file—it does not embed the *player* software. Playback of these animations depends on the appropriate software being present on the user's computer. Currently, Adobe Reader 6.0 used with the default installation of Microsoft Windows or Macintosh OS X will play animations in the AVI format with no additional media player software. Substantial changes in or the eventual obsolescence of the AVI format, such as its being dropped from future media player software, could mean that playback of the embedded animations would not be possible without manipulating the native data.

For the native data, there would be bit level preservation, but functional preservation strategies are undetermined. Emerging data exchange standards may make this task simpler, or it may be possible to identify CAD models of special interest and solicit software vendor support (free software, at a minimum) in migrating these models to that software product's most current version. However, technical capabilities change rapidly and a Preservation Policy Committee must be formed to periodically review and adjust preservation techniques. This Preservation Policy Committee should include representation from the registrar, the archival or curatorial department in charge of the collection and the information technology department.

[4] See: Stephen L. Abrams and David Seaman, Towards a global digital format registry, World Library and Information Congress, August 2003, available from ***www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf***; accessed 4 March 2004.

Accessing Digital Design Data

Figure 2.1f: Collection and Archiving System: Access and Dissemination Information Package (DIP)



The Access module enables searching of the archives using descriptive metadata such as project name, architect or date, and delivers the Dissemination Information Package. The Dissemination Information Package (DIP) is what is received by an end user searching the archives or a curator designing an exhibit. The DIP will contain the digital design data of interest, the associated metadata, and in some cases, the means for viewing or interacting with the data.

The dissemination responsibilities will likely be shared by the institution's collection management system (CITI for The Art Institute of Chicago) and the data repository system (DSpace). The collection management system will serve as the primary internal and public user interface for search and retrieval of information and will handle access controls. The data repository will provide for OAI-compliant (Open Archives Initiative) discovery by researchers and will provide for the delivery of the DIP to all users.

Role of Art Institute Collection Management System CITI

CITI (Collection Images Text and Index), The Art Institute's current digital cataloging system, uses the CDWA (Categories for the Description of Works of Art) metadata scheme for search and retrieval of data. Because CDWA was specifically designed for describing works of art and architecture, it has become the preferred metadata scheme for this purpose. This study has identified extensions to CITI that are consistent with the CDWA schema and would enable CITI to include all of the information fields that the Department of Architecture curators feel is relevant to their collections.

Archives conforming to the Open Archives Initiative (OAI) use a Dublin Core metadata scheme for search and discovery of their data. If The Art Institute chooses to join the DSpace Federation or be recognized as an OAI-compliant repository and make its digital collections widely available, it can create a programmed link so that CITItransfers the appropriate CDWA metadata fields down to Dublin Core fields in the DSpace repository.

Role of the Data Repository

In addition to storing and ensuring the preservation of the digital data, the data repository will receive requests from CITI and package the requested digital objects for dissemination. The most important element of dissemination for Web access is viewing the information.

Controlling Copyrights

Access to different forms of native data and different output resolutions for images will be controlled by two means. Anyone accessing the repository via a public interface, such as The Art Institute Web site, will be able to search for and discover all available materials, but they will automatically be restricted to retrieving only low-resolution images. They will then be able to request access to the additional material, subject to museum policies.

Viewing the Two-Tiered Collection

For viewing the output tier of the digital collection, low-resolution JPGs for browser-based viewing should be derived from high-resolution TIFF images during Ingest. PDF documents can be viewed by downloading a free PDF viewer.

For the second tier, consisting of native data including CAD files, acquiring (preferably through donation) and retaining copies of the original software used to create the native data in the collection would provide a means to view that data in its original form. These original programs would not, in most cases, allow Web-based viewing. Maintaining the hardware needed to run the software could also become a considerable burden, both because of the rapidity with which hardware becomes obsolete and unavailable and because of the inevitable—and potentially irreparable—hardware failures that come with age and use.

Another partial solution is to use proprietary multi-format viewers. These provide viewing access to a good range of formats, but not all. Also, as formats become obsolete, they may no longer be supported by commercial viewing software.

Viewers could be provided as Web server-based applications, although there would be a licensing cost associated with this option. Examples of current viewers are Brava!Viewer by Informative Graphics, Autovue by Cimmetry Systems, ViewCafe by Spicer Corporation and Roamer[3] by NavisWorks[3] (which is discussed below).

Although there is currently no comprehensive solution for making all native digital design data Web-accessible, this topic is of great commercial interest and better tools can be anticipated. The balance of this chapter discusses a range of currently available products for 2D viewing, 3D viewing and translation/repurposing.

2D Viewers

2D viewers allow users to view, dimension, and markup 2D CAD drawings without having the proprietary software in which the models were created. A 2D viewer is typically used to provide access to drawings in an intra-office or multi-office team setting. With online access to and viewing of drawings, different project participants can take off dimensions, markup the drawings and associate questions with the drawings. The more advanced 2D viewers will also print to scale with line weights supported.

Most 2D viewers accept files from major CAD systems, such as MicroStation and AutoCAD, as well as 2D images such as TIFF and PDF. Table 2.14 compares various 2D viewers with the graphics formats used by the design firms surveyed as part of *Section 1: Current State of Digital Design Tools and Data*. In addition, viewers often provide access to files in non-graphic formats.

3D Viewers

3D viewers allow users to view, navigate (or move around or through), measure and markup 3D CAD models without having the proprietary software in which they were created. Uses of 3D viewers follow the pattern of 2D viewers as they are often installed as a component of a local Document Management System or Web-based collaboration system. Markups are stored with author and date.

The more advanced 3D viewers allow the user to cut sections, view individual components or levels, explode (or break apart) the model, and view in shaded or wireframe modes.

Navisworks

Navisworks[3] Roamer[3] is an example of a 3D viewer with added functionality. The Roamer[3] opens a range of native file formats shown in Table 2.14 with all lighting and materials information and allows the user to navigate by zooming, rotating the model about an axis, orbiting around the focal point of the model or flying through a model. The user may also cross-section, measure and markup a model. An added functionality is the ability to create saved views and walk-through animations of the model.

The Navisworks[3] Publisher[3] plug-in to Roamer[3] adds the functionality of enabling the user to open many file types at once and publish to the compressed Navisworks[3] NWD format. Navisworks[3] offers a 3D viewer called Freedom to view the proprietary NWD format. Navisworks[3] provides Application Programming Interface to allow users to adapt and integrate its functionality to fit the user's programs.

Table 2.14: Graphics Formats supported by 2D and 3D Viewers

Design Products and Formats	Brava! Viewer (Informative Graphics) bravaviewer.com	MYRIAD (Informative Graphics) myriadviewer.com	Autovue (Cimmetry Systems, Inc.) cimmetry.com	ViewCafé Java Viewer (Spicer Corp.) spicer.com	Roamer[3] (Navisworks) navisworks.com
CAD, rendering & animation					
3ds Max (.max)					
ABVENT Art·lantis (.atl)					
AccuRender (.dwg)	x	x	x	x	
Adobe After Effects					
Adobe Premiere (.prproj)					
Alias Power Animator					
Alias Studio (.wire)					
ARRIS CAD					
AutoCAD (.dwg, .dxf)	x	x	x	x	x
Autodesk Architectural Desktop (.dwg)	x	x	x	x	x
Autodesk Architectural Studio (.asw)					
Autodesk Inventor (.idw, .ipt, .iam)		x			x
Autodesk MAP (.dwg)	x	x	x	x	
Autodesk Revit (.rvt)					
Autodesk VIZ (.max)					x
Camtasia Recorder (.avi)					
DESTINI V1.0					
DOE-2					
Electric Image Universe					
formZ (.fmz)					
Graphisoft ArchiCAD (.pln)					
Graphite (.vc6)					
Lightwave 3D (.lw, .lwo, .lws)					
Macromedia Director (.dir, .dcr)					
Maya (.ma, .mb)					
Media 100 (.mov)					
MicroStation (.dgn)	x	x	x	x	x
MicroStation TriForma (.dgn)	x	x	x	x	x
Nemetchek Allplan (.000)					
NuGraf (.bdf)					
Pro/Engineer (.asm, .prt)		x	x		
Rhinoceros (.3dm)					
SketchUp (.skp)					
SolidWorks (.slddrw, .sldprt, .sldasm)	x	x	x		x
Surfware SURFCAM (.dsn)					
VectorWorks (.mcd)					
Image-manipulation, presentation & page layout					
Adobe Acrobat (.pdf)	x	x	x	x	
Adobe Acrobat Distiller (.ps)	x	x	x	x	
Adobe Illustrator (.ai)		x	x		
Adobe InDesign (.indd)					
Adobe Pagemaker (.pmd)					
Adobe Photoshop (.psd)		x			
Alias Sketchbook Pro (.jpg, .tiff)	x	x	x	x	
AlphaGraphics E-Z Pix (.jpg, .tiff)	x	x	x	x	
CorelDRAW (.cdr)	x	x	x		
Macromedia Flash (.fla, .swf)					
Macromedia Freehand (.fh10)					
Microsoft PowerPoint (.ppt)	x	x	x	x	
QuarkXPress (.qxd)					
Non-graphic design support					
ARCHIBUS (.dbf – Oracle database)					
Filemaker (.fmp, .fpt)					
Microsoft Access (.mdb)			x		
Microsoft Excel (.xls)	x	x	x	x	
Microsoft Word (.doc)	x	x	x	x	
Neutral formats					
IGES	x	x	x	x	x
STEP		x	x		x

3D Collaboration Tools

3D Collaboration Tools take the exchange of 3D data to the next level. The most advanced collaboration tools allow for edits to be made to the model, something the 3D viewers do not. Since they allow editing of the original data, how useful such tools would be to the Department of Architecture is a question. They might permit interesting interactions with archived models by students or researchers.

These collaboration tools are sometimes associated with a single proprietary CAD system that the host party is required to have. In some cases, one proprietary license can be shared with all members logged into an online meeting.

So far, these tools address the needs of manufacturers rather than architects. Current systems are designed for the data formats and object types found in mechanical CAD systems.

Repurposers

Repurposing software has the capability to import a CAD model in one file format and then export it to a different file format or presentation format, such as a navigable 3D model view for a PowerPoint presentation. Some repurposers also include a repository for archiving the native data files. These products are of some interest because they combine the functionality of a repository with a viewing and repurposing tool. However, they are proprietary, rather than open source, which makes them poor candidates for a long-term digital archive solution. Several programs are discussed here as examples of what is currently available.

Right Hemisphere

Right Hemisphere has created a product called DEEP SERVER that archives, searches, views, translates, animates, and publishes 2D and 3D CAD data in a range of formats.

DEEP SERVER captures more information than just geometry, such as layer information, lights, surface materials, and cameras. It has plug-ins that also capture saved views created in AutoCAD. DEEP SERVER also is capable of translating CAD data from one format to another. (See Table 2.15 for import and export formats.)

DEEP SERVER can be configured to have the following default outputs:

- Line drawing in encapsulated Postscript format
- Large-scale, high-resolution image for exhibiting in TIFF format
- Small-scale, high-resolution image for publication TIFF format
- Small-scale, low-resolution image for PowerPoint presentation in JPG format.

The user can also embed a *navigable* 3D model in Microsoft PowerPoint, Word, and HTML (with PDF format promised in the future). Another Right Hemisphere product, DEEP PUBLISH, gives an even greater range of publishing options.

DEEP SERVER is designed to run on PCs, though a Viewpoint renderer can be installed on the server to make the information available to Mac users.

UGS

Teamcenter Solutions, created by UGS, is a Web-based or Windows-compliant data repository that employs UGS's VisView product to view a range of CAD formats and publish to image formats. VisView is used as a collaboration tool by Boeing, GM, Ford and Honeywell.

To view models in Teamcenter Solutions, UGS uses VisView, a 2D/3D viewer that allows navigation, layer and object management, measuring, sectioning and mark-ups. VisView accepts a variety of file formats and requires translating modules to do so. VisView renders 3D files in its own neutral format, JT, and includes only geometry and color information. The user can export in the native file format or publish to HTML, JPG or TIFF formats. With the addition of Vis Concept, the user can publish to presentation formats that project 3D objects in a virtual reality CAVE (Cave Automatic Virtual Environment).

VisView does not have the capability to translate CAD data from one CAD format to another, nor does it capture information beyond the geometry and color of the model, such as material attributes, lighting or previously saved views.

Okino Computer Graphics

Okino Computer Graphics has developed a repurposer called the PolyTrans 3D Translation System that allows users to import from one 3D CAD package, view and manipulate the model information and export to a different 3D CAD package or other presentation format. It offers a wide range of input CAD formats and is good at capturing all visualization information in the model. Its strength is high-quality renderings and presentation and publication formats. (See Table 2.15 for a comparison of accepted input and output formats.)

Table 2.15: Accepted Formats for Selected Repurposers / Translators

Design Products and Formats	**DEEP SERVER by RIGHT HEMISPHERE** www.righthemisphere.com		**PolyTrans 3D Translation System by Okino Computer Graphics** www.okino.com	
CAD, rendering & animation	Import	Export	Import	Export
3ds Max (.max)	x	x	x	x
ABVENT Art·lantis (.atl)				
AccuRender (.dwg)	x			
Adobe After Effects				
Adobe Premiere (.prproj)				
Alias Power Animator				
Alias Studio (.wire)				
ARRIS CAD				
AutoCAD (.dwg, .dxf)	x		x (.dxf only)	x (.dxf only)
Autodesk Architectural Desktop (.dwg)	x			
Autodesk Architectural Studio (.asw)				
Autodesk Inventor (.idw, .ipt, .iam)			x	
Autodesk MAP (.dwg)				
Autodesk Revit (.rvt)				
Autodesk VIZ (.max)	x	x	x	x
Camtasia Recorder (.avi)	x	x		
DESTINI V1.0				
DOE-2				
Electric Image Universe				
formZ (.fmz)				
Graphisoft ArchiCAD (.pln)				
Graphite (.vc6)				
Lightwave 3D (.lw, .lwo, .lws)	x	x	x	x
Macromedia Director (.dir, .dcr)				
Maya (.ma, .mb)	x	x	x	
Media 100 (.mov)				
MicroStation (.dgn)				
MicroStation TriForma (.dgn)				
Nemetchek Allplan (.000)				
NuGraf (.bdf)			x	x
Pro/Engineer (.asm, .prt)			x (render .slp only)	x (render .slp only)
Rhinoceros (.3dm)	x		x	x
SketchUp (.skp)				
Solidworks (.slddrw, .sldprt, .sldasm)	x		x	
Surfware SURFCAM (.dsn)				
VectorWorks (.mcd)				
Image-manipulation, presentation & page layout				
Adobe Acrobat (.pdf)				
Adobe Acrobat Distiller (.ps)				
Adobe Illustrator (.ai)			x	
Adobe InDesign (.indd)				
Adobe Pagemaker (.pmd)				
Adobe Photoshop (.psd)	x		x	x
Alias Sketchbook Pro (.jpg, .tiff)	x	x		
AlphaGraphics E-Z Pix (.jpg, .tiff)	x	x		
CorelDRAW (.cdr)				
Macromedia Flash (.fla, .swf)			x	
Macromedia Freehand (.fh10)				
Microsoft PowerPoint (.ppt)		x		
QuarkXPress (.qxd)				
Non-graphic design support				
ARCHIBUS (.dbf – Oracle database)				
Filemaker (.fmp, .fpt)				
Microsoft Access (.mdb)				
Microsoft Excel (.xls)				
Microsoft Word (.doc)		x		
Neutral formats				
IGES	x		x	x
STEP			x	x