

Preservation and Access for Born-digital Electronic Records: The Case for an Institutional Digital Content Format Registry

Kyle R. Rimkus, Bethany Anderson, Karl E. Germeck, Cameron C. Nielsen, Christopher J. Prom, and Tracy Popp

ABSTRACT

Since 2014, the University of Illinois at Urbana-Champaign Library has taken custody of a growing number of collections of “born-digital” records, largely through the University Archives. These collections comprise a panoply of digital content formats, ranging from those in common use to obscure varieties from the early days of personal computing. As such, they pose a challenge to digital preservation and access. Knowing what software to use to open files in formats that have fallen out of use is often difficult, let alone installing obsolete software on contemporary operating systems. At the same time, the sheer bulk of collections, as well as an accelerating rate of born-digital accessions from faculty and campus offices, makes it difficult to assess these files at the time of acquisition. These challenges suggest the need for preservation policies on digital formats in collections of electronic records, as well as for firsthand knowledge of the software required to facilitate curator control over and patron access to these collections. This article presents an overview of an evolving approach taken by archivists and librarians at the University of Illinois at Urbana-Champaign to build the policies, technical knowledge, and systems for an effective preservation and access program for electronic records. Their implementation of a local digital content format registry, while young, suggests that archivists and digital preservationists would benefit from further development of tools and practices focused on born-digital formats, and the thoughtful integration of institutional knowledge with international format registries.

© Kyle R. Rimkus, Bethany Anderson, Karl E. Germeck, Cameron C. Nielsen, Christopher J. Prom, and Tracy Popp.



KEY WORDS

Digital preservation, Born-digital, File formats, Digital archiving, Appraisal

Over the past twenty years, the field of digital preservation has seen an evolution in thinking about the importance of file format policy for repositories. In the 1990s, much writing on the subject was speculative in nature, with authors like Donald Waters and John Garrett forecasting that digital repository managers might, to confront the challenge posed by large collections of disparate materials, “migrate digital objects from the great multiplicity of formats used to create digital materials to a smaller, more manageable number of standard formats.”¹ This approach was countered most notably by thinkers such as Jeff Rothenberg, who argued instead for emulating the software environments of obsolete file formats—that is, providing software to mimic as closely as possible the computing environment in which the files were originally created and used—as the surest path to providing faithful access to their content.² Rothenberg’s advocacy of emulation was not without its detractors,³ and many digital preservation professionals who attended digital library conferences in the aughts will recall lively debate among those who favored “normalization” strategies based around trusted file formats versus advocates of software emulation.⁴

With regard to file format policy research, several studies from 2005 through 2008 (e.g., work led by the National Library of the Netherlands,⁵ the National Library of Australia,⁶ Stanford University,⁷ and the Online Computer Library Center⁸) sought to identify risk factors inherent to file formats and to define what makes a “good” file format for digital preservation. Numerous cultural memory organizations complemented these efforts by publishing policies of trusted or preferred file formats for long-term stewardship. It is common to hear digital preservation managers proclaim the tagged image file format (TIFF) as their trusted image file format, or the waveform audio file format (WAV) as their preferred audio file format, without controversy.

A 2013 publication surveying and analyzing such file format policies at Association of Research Libraries (ARL) member institutions found that most of these policies are “very much rooted in relatively small-scale data management practices—stewarding files through digitization workflows, for example, or curating a university’s research publications,” but that “[a]s libraries and archives begin to set their sights on collections of heterogeneous files such as born-digital electronic records and research data, this is expected to spur on further evolution not only in the file formats that appear in digital preservation policies, but in the way file format policies are articulated and implemented.”⁹ This finding underscores the assertion that digital preservation professionals in libraries and archives tend to be most comfortable with file formats that result from digitization efforts, or workflows whose end result could be said to stand

in for traditional physical media, in other words digital surrogates representing printed pages, phonographic recordings, or moving image films. On the other hand, archivists do not possess reliable tools for stewarding structures native and unique to the networked digital realm, such as hyperlinked, interactive, or editable information content, including everything from flash games embedded in web pages authored in the hypertext markup language (HTML) and cascading style sheets (CSS) to specialized three-dimensional models. Most of these materials are unsuitable for mass migration to trusted file formats due to the technical and legal hurdles involved, which calls for new thinking about what exactly to preserve, as well as how.¹⁰

Even for fairly straightforward content types such as those generated by the digitization of physical media, the field is pivoting away from rigid conceptions of what constitutes a trusted file format. Kevin DeVorsey and Peter McKinney, in a study of file format risk at the National Library of New Zealand, concluded that “files contain multifarious properties. These are based on the world of possibilities that the format standard describes, but can also include non-standard properties. The range of possibilities and relationships between them is such that it is quite meaningless to purely measure a file’s adherence to the format standard.”¹¹ In other words, the common practice in libraries and archives of making a short list of trusted file formats is inadequate, because what any single file format may contain is, in most cases, highly variable. If preservation implies access to the content preserved, preservationists must possess knowledge of software environments and the dependencies necessary to create authentic renderings of digital bitstreams. That is, they need to open files and access the information they contain in an accurate form. Or, as the authors of InterPARES 1 found, “Empirically, it is not possible to preserve an electronic record: it is only possible to preserve the ability to reproduce the record.”¹²

At one level, rendering capability should be an essential aim of digital preservation. In practice, however, this has proven difficult to implement, due to what some see as flawed metaphors of digital “objects” and “records” imported from traditional preservation work. These metaphors, Christoph Becker argues, obscure the true nature of files as components within a software system, the interactions of which “produce emergent properties that we cannot attribute to the parts, only to the whole.”¹³ Any consideration of what constitutes significant properties of acceptable renderings of preserved files must account for inputs from all parts of the operating environment (fonts, color profiles, dependencies, and the like). Observing that digital file “damage occurs often not as a loss of physical integrity, but as a loss of relationships between elements, whether through link rot, obsolescence, or lack of metadata,” Becker sees strategies based solely on migrating files from one format to another as fundamentally inadequate to ensuring the full accessibility of preserved digital materials

for posterity. Rather, Becker argues that the key to preserving meaningful rendering capability along with individual units of encoded content is preserving knowledge of the critical relationships between the constituent components of software environments.

Considering the importance of software relationships and rendering capability in meaningfully preserving digital content, it is notable that such information is largely absent from those file format registries that digital preservationists rely on most as technical reference tools. While PRONOM, an internationally recognized registry of file format information managed by the National Archives of the United Kingdom, does include a metadata field for “technical environment,” this field is in most cases unpopulated. In addition, other efforts to create rich resources in the realm of file format and preservation policy description, such as National and State Libraries of Australasia’s Digital Preservation Technical Registry,¹⁴ the Preservation Actions Registries project led by the Open Preservation Foundation,¹⁵ or the Scaling Emulation as a Service Infrastructure¹⁶ (EaaS) project’s *Software Metadata Recommended Format Guide*, have yet to broadly share their findings.

Perhaps archivists and preservationists would benefit by augmenting the universal information stored in international file format registries with local knowledge gained from hands-on experience with locally curated born-digital materials, locally available software, and locally available operating system environments. While this practice is not yet widespread, it has been attempted, most notably by the National Library of Australia, whose Digital Preservation Knowledge Base is the institution’s “first practical step to equipping the digital preservation unit with essential knowledge about the file formats present in the Library’s collections and their relationships with software applications.” The National Library of Australia is recording information about file formats and the software environments needed to access them in spreadsheets, with plans for migration to a database and a linked-data store intended to represent the library’s rendering capability in all its complexity.¹⁷

Based on the current state of the field, this article describes an effort by librarians and archivists at the University of Illinois at Urbana-Champaign to build a local digital content knowledge base using technology and a data model derived from institutional practice. We seek to add to a young but growing body of literature on the topic of developing best practices in organizational digital content format monitoring and the appraisal and processing of born-digital materials. Although we undertook this project to meet a local need for file format policy management, its results suggest that the community could benefit from new directions in digital content format research. Specifically, it suggests that digital preservationists should find ways to better integrate largely hidden local expertise and distributed knowledge with centralized format registries.

Background

The University Archives at the University of Illinois at Urbana-Champaign has been collecting born-digital content since the early 2000s, with early efforts centered on hybrid collections (holding both paper-based and electronic materials) and born-digital personal archives and student organization websites.¹⁸ These records were acquired on a variety of media, including optical disks, floppy disks, and external drives, as well as laptops and desktop computers. From 2008 to 2009, one of this article's authors began researching best practices for preserving born-digital content and developing a base of knowledge and recommendations for future services.¹⁹ In 2011, the library hired a digital preservation coordinator to develop infrastructure for the acquisition and appraisal of computer media who established a "born-digital" lab where concepts and technologies from the field of digital forensics informed hardware and software choices. Initially, archivists deferred the curation of these materials to a future date by consigning born-digital acquisitions to a folder titled "unprocessed" on a secure server maintained by the library, while staff tested appraisal and processing software and tools.²⁰ In 2014, however, staff in the library's Preservation Services and Information Technology units introduced a digital preservation repository service called Medusa²¹ to aid collection curators in the long-term stewardship of digital content and began developing practices to address the backlog of digital acquisitions.

Medusa runs on locally developed open-source²² software written in the Ruby on Rails web framework.²³ Its design was inspired by the Reference Model for an Open Archival Information System (OAIS),²⁴ a prominent standard for digital preservation service architecture. Medusa's use is presently limited to the University of Illinois at Urbana-Champaign Library, where its primary users are library and archives collection managers in repository units, with these defined as groups responsible for curatorial decisions related to the preservation of, access to, and rights status of collections of digital content (e.g., an archives, institutional repository, or departmental library). At present, Medusa's collections comprise born-digital books, manuscripts, photographs, audiovisual materials, scholarly publications, and research data from the library's special collections, general collections, and institutional repositories.²⁵

The University Archives has processed and ingested digital content at a rate of approximately 4 to 5 terabytes per year for the past five years and has deposited, as of June 2020, a total of 41 terabytes (approximately two million files) into Medusa.

It is also worth noting that the University of Illinois at Urbana-Champaign is home to a prominent iSchool and that graduate student assistants have

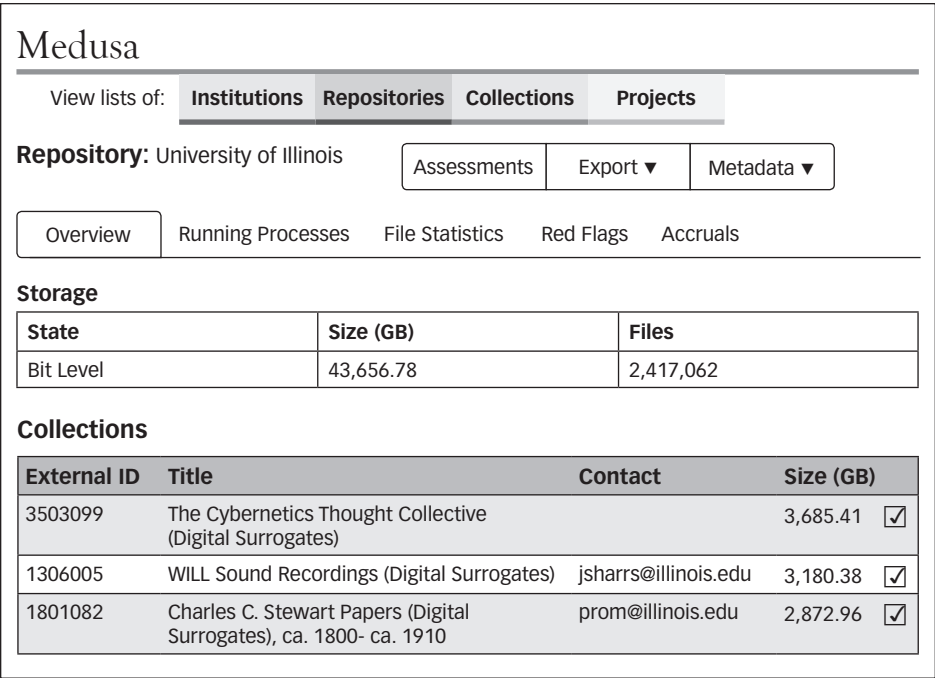


FIGURE 1. Homepage of the University Archives in Medusa

provided ongoing support for workflows related to the preservation of born-digital collections since the inception of the library’s digital preservation program. Indeed, the authors of this article comprise a team of professional librarians and archivists and a (then) graduate assistant, with additional research support provided by (former) graduate student workers Scott Witmer and Shreya Udhani.

Methodology

We identified a working set of born-digital files in the Medusa digital preservation repository. This data set was intended to consist entirely of born-digital materials and, as such, did not contain library-digitized items or collections that mix digitized content with other digital materials acquired from archival donors. To this end, programmer Howard Ding created a system feature that allowed curators to group collections of born-digital collections under the heading of a single “Virtual Repository.” Once complete, the “Born-digital Virtual Repository” permitted us to view our data set under a single dashboard that listed constituent collections and provided aggregate data on the total number of files, the total number of files with a given extension, and the total number of files of a given media type. At the time of writing this article, the Born-digital

Downloaded from https://prime-pdf-watemark.prime-prod.pubfactory.com/ at 2025-07-01 via free access

Collections Virtual Repository, which has served as a working set for this article, includes 151 collections comprising 1.35 million files, for a storage total of 2.4 terabytes. These collections contain a mix of administrative records, publications, and personal papers.

After isolating the data set described here, the authors worked again with repository programmer Howard Ding to establish a born-digital content format registry feature in Medusa to serve as an eventual reference source for policies and knowledge related to locally held digital content types.²⁶ Initially thought of as a “File Format Registry,” the authors adopted the term “Digital Content Format Registry” to underscore the difference between it and international file format reference resources²⁷ like PRONOM and to allow for multiple entries for different varieties of file formats that constitute distinct ways of packaging digital content. To clarify these distinctions, consider the JPEG-2000 file format, which has an entry in PRONOM.²⁸ Whereas PRONOM lists the attributes and qualities of JPEG-2000 as a format (and of all files that exist in this format), the authors recognized, based in part on prior research,²⁹ that collections in their custody comprise several different flavors of JPEG-2000, depending on what scripts created the original files, with each of these variants presenting different barriers to access. Thus, re-creating the work of cataloging the significant properties of the JPEG-2000 file format had no utility per se; the need existed, however, to document recurring local variants of the JPEG-2000 file type and the most commonly encountered challenges to accessing them.

Based on this thinking, the Digital Content Format Registry contains the following general fields for all entries:

Field	Description
Name	A descriptive, long-form name to identify a particular type of locally held digital content. While these names may simply reference a file format, the concept of a digital content format is flexible enough to encompass variations of file format content profiles, e.g., recurring instances of file formats that share common characteristics.
PRONOM ID	Relevant PRONOM unique identifiers (PUIDs) that link to corresponding file format entries within the PRONOM Registry Database. In this model, the digital content format entry is a local enhancement of a base file format entry from PRONOM, to document characteristics of digital content formats under institutional stewardship.
Policy Summary	A statement describing local confidence level and curatorial practice with regard to a digital content format, if this exists
Logical [File] Extension	A list of known file extensions native to the file format or to the program/application from which it originated.
Related Formats	A list of related digital content formats from within the Digital Content Format Registry, some of which may share consistent dependencies.

In addition to these general fields, digital content format entries also contain sections focused on file-rendering profiles and normalization paths, as well as Administrative Notes and Attachments. File-rendering profiles detail the software and operating systems used to open files, and normalization paths specify information about target file formats to which designated formats may be migrated should the need arise, with information about recommended software for conversion.

With the format registry feature in place in Medusa, the authors began creating records in it. While we often identified and documented file formats in a nonlinear fashion, we generally followed these steps for each digital content format:

1. **Selection.** From Medusa’s Virtual Repository of Born-digital Collections, a project researcher navigated to the File Statistics tab and selected an extension from the File Extension table to investigate (see Figure 2).
2. **Investigation.** The researcher examined born-digital files associated with the selected extension, and attempted to identify the file format and to access the file’s content by using a variety of tools, techniques, and resources. The approaches used differed from one format to another, but often included examining Medusa’s built-in file identification data such as a file’s assigned internet media (MIME) type or technical information provided by the file information tool set (FITS), looking at selected files in a text/hex editor, culling information from trusted online file identification tools and file format registries, and drawing contextual information from the collection’s directory and

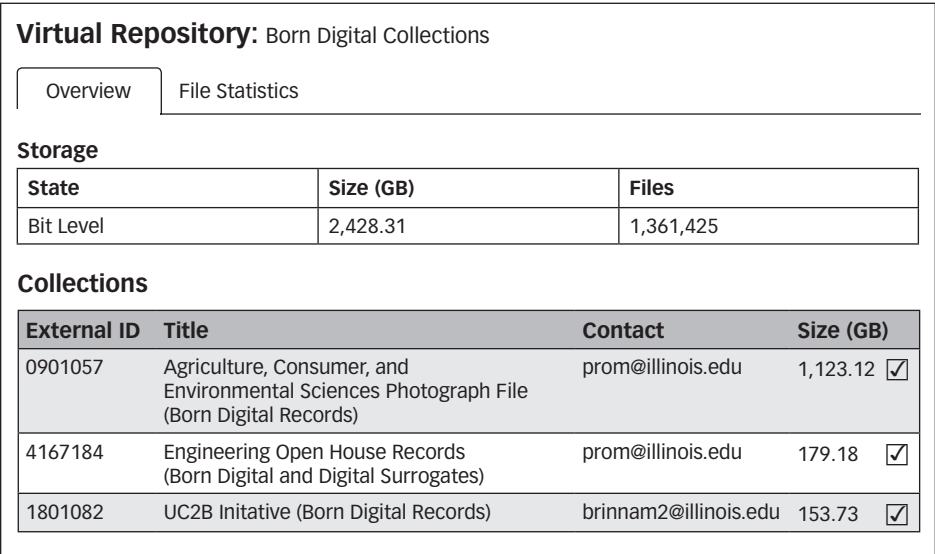


FIGURE 2. Partial list of collections in the Born-digital Collections Virtual Repository

subdirectory hierarchies, as well as from other nearby files in which the files under investigation appear. Then, if applicable or possible, the researcher opened selected files in their native and/or a compatible OS and software environment. This investigation sometimes required contacting creators of the collection; vetting and synthesizing available information from specialized blogs, forums, and listservs; and downloading and experimenting with software.

3. **Synthesis.** The researcher created a Digital Content Format Registry entry or entries. At minimum, this included a descriptive file format name, any associated PRONOM IDs (if applicable), all known file extensions native to the format, related file formats within the registry, and, if enough information was available, at least one file-rendering profile detailing the software and operating systems needed to successfully access the digital content in question. For each selected file format entry, the researcher provided contextually rich notes and any helpful online and/or local documentation, such as detailed information about relationships and dependencies with other related file formats within the registry, and, if applicable, a policy summary to reflect local confidence in the file format, different types of descriptive notes, additional rendering environments, and normalization paths.

The Research Process

During a two-year period from September 2017 to September 2019, the research team created entries for 250 digital content formats in Medusa's publicly accessible Digital Content Format Registry.³⁰ This iterative process involved revising registry entries on an ongoing basis, re-identifying formats that had been originally misidentified, or identifying additional extensions belonging to a format that were not included within the original registry entry. These entries range from brief descriptions containing a short one- or two-line phrase describing the content format in an administrative note, all the way to rich entries featuring three or more administrative notes that weave together local contextual information and relevant information from online and book resources about the format and/or about the files in Medusa that belong to the format. All entries, regardless of their description level, have, at minimum, a content name and logical extension(s). Depending on available information, entries may also contain relevant PRONOM IDs, a list of related formats, rendering profile(s), normalization paths, attachments, and a policy summary.

Looking at file statistics alone in the Born-digital Virtual Repository is illuminating, especially compared with other repository units in Medusa. The Map Library, for example, houses 82,000 files of what could be considered digital

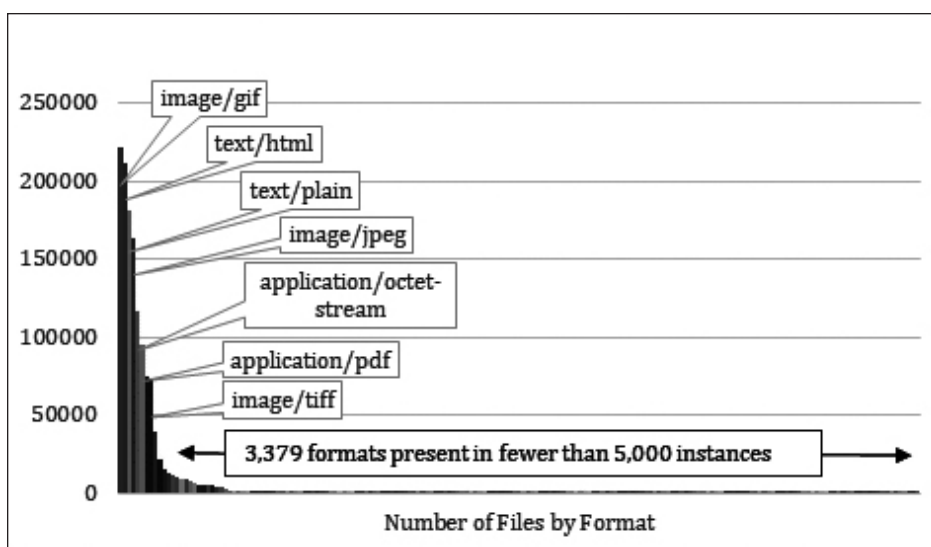


FIGURE 3. The project data set's 1.3 million files by format

surrogates of paper maps. These files comprise eight MIME types, of which the vast majority are image/tiff, text/xml, image/jp2. By contrast, the Born-digital Virtual Repository's 1.3 million files includes 184 distinct MIME types with a broad distribution over different varieties. That is, 37 of these MIME types are present in numbers greater than 1,000 files, 39 MIME types count between 10 and 1,000 files, and 107 MIME types have fewer than 100 files. As Figure 3 shows, these numbers demonstrate a “head” of a few formats in large number, followed by a “long tail” of many formats present in much smaller numbers.

When confronted by such a forest of preservation issues, knowing how to begin approaching its individual trees is difficult. The research process is best demonstrated, however, by a detailed description of the steps taken to flesh out a comprehensive set of entries for a particular digital content format, in this case the Macromedia Director protected movie format. While the deep dive that follows may go too far into technical detail for some readers, it will illustrate just how much effort is sometimes required for archivists and preservationists to understand and gain access to the digital materials under their control. In addition, it will open the door to an analysis of how best to integrate this effort into local practices for appraisal, processing, and long-term access and preservation.

Researching a Format

The following walkthrough demonstrates the steps co-author and project researcher Karl Germeck took to select and identify a born-digital format, test the rendering of representative files, and document knowledge gained about

that format in an entry or set of entries in the Medusa Digital Content Format Registry. Germeck began by browsing the File Statistics tab for Medusa's Virtual Repository and selecting extensions at random for research. He then noted ninety-five files with the unfamiliar extension "DXR" and selected them for further investigation. A list of these files revealed that they were predominantly found in the same collection, the Stanley Smith Papers (Born-digital Records, Digital Surrogates and Audiovisuals).³¹

For some background, the Stanley Smith Papers (Born-digital Records, Digital Surrogates and Audiovisuals) comprise an entirely born-digital collection featuring personal records from a professor of chemistry and chemical education (1960–2010) at the university. They contain presentations, images, web tutorials, computer programs, software code, and audio/visual materials concerning chemistry curricula and instruction, with topics including web-based instruction, educational software, course materials, and Chemistry Department equipment and facilities. They are an excellent example of the challenges archivists encounter in stewarding born-digital records.

The files selected with extension DXR were created in 1995 using Macromedia/Adobe Director, a multimedia authoring and internet publishing tool for Shockwave, a web-based application and video game platform (not to be confused with the similar web-based platform, Adobe Shockwave Flash).³² Director achieved commercial popularity during the 1990s and has been used to create two- and three-dimensional animation sequences, online video games, educational software applications, self-running interactive kiosks, stand-alone CD-ROMs and DVDs for Windows and Mac environments, as well as applications for Apple mobile devices. Four of Director's primary file types include an uncompressed movie (DIR), a compressed movie (DCR), a protected movie (DXR), and a Windows/Macintosh projector (more on projector files to follow). In 2017, Adobe announced that it would no longer support the development of Director or the Shockwave platform in favor of its emerging Creative Cloud technology.³³

While this information is readily available on the Web, archivists who seek to preserve and provide access to content generated by these obsolete software tools must go a step further. The project researcher who selected DXR files for investigation sought to verify that the DXR files in custody were what they purported to be, that is, that they were not in fact another file format that shares the DXR extension. In the case of the DXR content isolated for study, a detailed FITS identification report on a single file entitled ACCL_ROH.DXR (see Figure 4) indicated that the digital record and object identification (DROID) tool identified the format as "Macromedia Director" and its MIME type as "application/x-director." To further confirm the identity of the file in question, the researcher opened it in Notepad++,³⁴ a text editor with a hex editor plugin. In Notepad++, he was able to view the file's raw bitstream content and to compare this to

technical information on the Macromedia Director format from the United Kingdom National Archives’ PRONOM database. The opening bit-sequence of a Director for Windows file, represented in ASCII as RIFX, matched the bit-level examination of the opening sequence of the test file shown in Figure 5.³⁵ To bolster this evidence, the researcher then viewed the file as plain text to scan for embedded metadata of interest in the file headers. Indeed, a key word search of the file for the word “director,” displayed in Figure 6, also revealed the text “Director 4.0,” suggesting that Macromedia Director version 4.0 was used to create the file.

Having verified the file format and learned more about it by gleaning contextual information from Adobe’s website and from Wikipedia entries related to

```
▼<fits xmlns="http://hul.harvard.edu/ois/xml/ns/fits/fits_output"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output
http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="0.10.1"
  timestamp="10/21/16 1:50 PM">
  ▼<identification status="SINGLE_RESULT">
    ▼<identity format="Macromedia Director" mimetype="application/x-director"
      toolname="FITS" toolversion="0.10.1">
      <tool toolname="Droid" toolversion="6.1.5"/>
      <version toolname="Droid" toolversion="6.1.5">PC</version>
      <externalIdentifier toolname="Droid" toolversion="6.1.5" type="puid">x-
        fmt/341</externalIdentifier>
      </identity>
    </identification>
```

FIGURE 4. FITS file identification results for file ACCL_ROH.DXR

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded text
00000000	52	49	46	58	00	02	B6	FC	4D	56	39	33	69	6D	61	70	RIFX...µMV93imap
00000010	00	00	00	18	00	00	00	01	00	00	00	2C	00	00	00	00
00000020	00	00	00	00	00	00	00	00	00	00	00	00	6D	6D	61	70mmap

FIGURE 5. File ACCL_ROH.DXR viewed in a hex editor

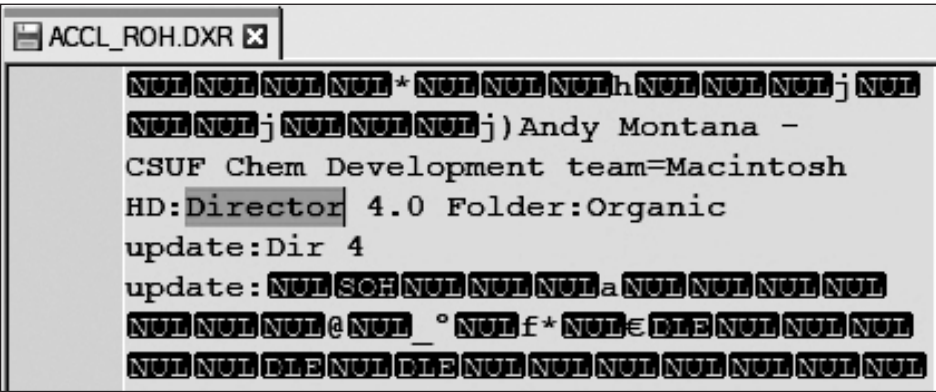


FIGURE 6. File ACCL_ROH.DXR opened in a plain text editor

Adobe Director and Adobe Shockwave, the researcher then made several unsuccessful attempts to open and view the ACCL_ROH.DXR test file in the Windows 10 environment. An initial attempt to open the file within Adobe Director 12.0 (the last supported version of Director for Windows) failed because the file was “protected and cannot be opened.” A second attempt to open ACCRL_ROH.DXR in the Firefox web browser enabled with the most current version of the Adobe Shockwave Player plug-in produced a static graphic displaying an equation titled Acid Chloride + Alcohol, accompanied by a Shockwave message indicating errors with the file prevented playback. Interestingly, both of these testing tools are obsolescent, as the latest version of Firefox no longer supports the Director Shockwave plugin, and Adobe no longer provides a download for the trial version of Director 12.

Following these two failed attempts to sufficiently open ACCL_ROH.DXR, the researcher consulted the book *Macromedia Director Lingo Workshop* in an effort to better understand the specific nature of Director-protected movies. Written by John Henry Thompson, chief engineering scientist at Macromedia from 1987 to 2001 and the inventor of Lingo, the scripting language that powered Director and the Shockwave platform, this book revealed that in the Director publishing process, DXR files share a dependency with a related application.

Specifically, there had been two ways to publish and distribute Director content: 1) as web-based Shockwave movies, in which case the published movies hold the DCR extension, or 2) as a Windows or Macintosh run-time environment known as a “projector”—often taking the form of a software program or application—and distributed onto removable media such as floppy disk, CD-ROM, DVD, or Apple mobile device. Within Windows, a Director projector held the EXE extension.³⁶ When preparing Director movies for distribution on removable media, content creators had the option of protecting their DIR uncompressed movie files via encryption as a means of safeguarding the intellectual property of their work. If the creator chose the “protect movie” option, the software encrypted and compressed the video stream of the originating DIR file and changed the file’s extension from DIR to DXR. Once encrypted, a Director movie that had been “protected” could no longer be manipulated or rendered within Director.³⁷ Finally, when a projector containing Director content was created for distribution on removable media, its creator could either compile video files within the projector itself or link the video files to the projector. If linked, the software often placed originating DIR or DXR files within the main directory or a subdirectory where the projector was located.³⁸

This was the case with ACCRL_ROH.DXR. As a Director-protected movie, it was encrypted and altered so that one could no longer open it in Director, or play it in Adobe Shockwave Player,³⁹ without access to its dependent projector file. Indeed, closer examination of the file directory structure of the Stanley

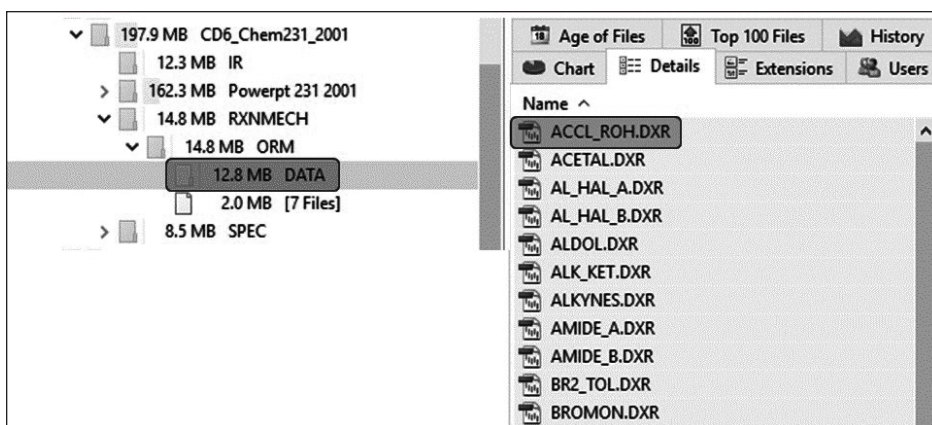


FIGURE 7. Structure and contents of DATA folder, with ACCL_ROH.DXR highlighted

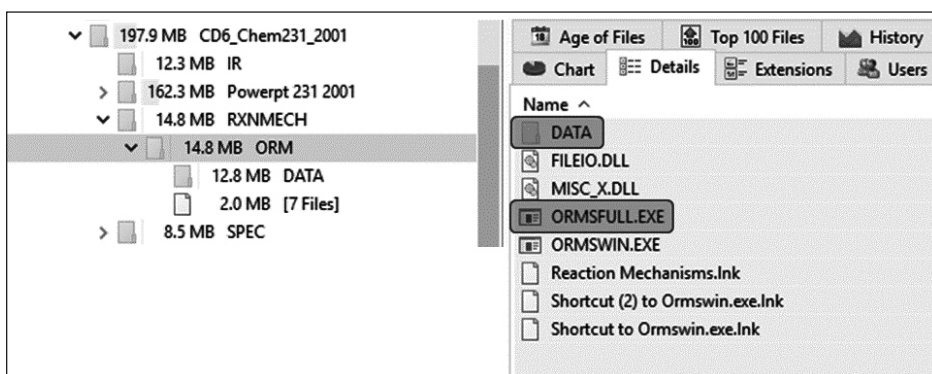


FIGURE 8. Structure and contents of the parent ORM folder, with the executable file ORMSFULL.EXE and the DATA folder

Smith Papers revealed that the DATA directory containing ACCRL_ROH.DXR (and a number of other DXR files) sits within a parent directory named ORM, which contains the executable file ORMSFULL.EXE (see Figures 7 and 8). This realization underscored that born-digital files cannot, and should not, be assessed as solitary objects divorced from their larger documentary context and that idiosyncratic file types may appear to cause errors when the user (or the archivist) does not fully understand their required dependencies.⁴⁰

The researcher then conducted a bit-level investigation of the file ORMSFULL.EXE. By cross-checking file signature values revealed in a hex editor (see Figure 9) with file signature information listed in PRONOM, the researcher identified the file as a 16-bit New Executable, a file format compatible with Windows 3x, Windows 95, Windows 98, and Windows ME 16-bit OS platforms.⁴¹ In addition, the two following pieces of readable text drawn from viewing ORMSFULL.EXE within a hex editor further suggested the file's relationship to the Windows (rather than the MS-DOS) platform and Macromedia Director: 1) a piece of DOS

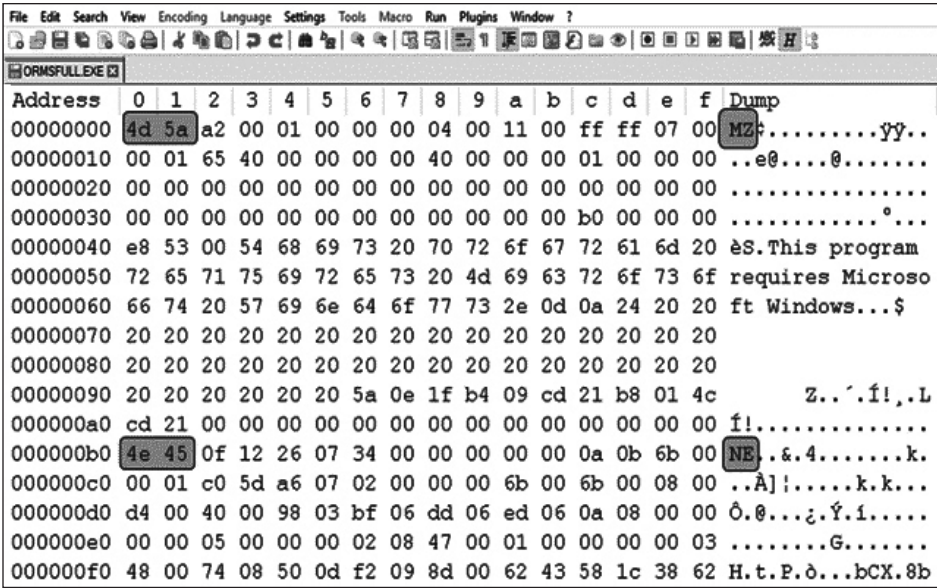


FIGURE 9. Viewing ORMSFULL.EXE in a hex editor

warning code—known as a “stub” and used to indicate that the executable will not run in DOS—reads, “This program requires Microsoft Windows,” and 2) the text “Director for Windows Release 4.0.4.”

Based upon this textual evidence discovered in embedded file header meta-data, as well as previous knowledge about the OS requirements for Macromedia Director 4.0,⁴² the researcher determined that ORMSFULL.EXE was originally intended to run on Windows 3.1. To test this conclusion, the researcher installed an instance of Windows 3.1 in a DOSBox emulator, a program designed to approximate the user experience of the DOS operating system. The researcher then transferred a copy of the ORM directory containing ORMSFULL.EXE and the subdirectory of DXR files to the Windows 3.1 file system and placed a shortcut to ORMSFULL.EXE on the Windows 3.1 desktop.

Upon launch of ORMSFULL.EXE, two animated, boulder-sized molecules lumbered across the screen and smashed together to reveal the title frame of a commercially purchased edition of the interactive chemistry education software, Organic Reaction Mechanisms, School Edition (ORMS) 2.0.⁴³ The program’s main menu displayed a series of chemical reaction categories in the upper portion of the screen (see Figure 10). When one of the reaction categories was selected, a set of animated chemical reaction modules appeared in the lower-left portion of the screen (see Figure 11), with each individual reaction module corresponding to one of the DXR files stored in the DATA folder within the ORM directory.⁴⁴ The acid halide, rx with ROH module, for instance, corresponded to ACCRL_ROH.DXR and consisted of four animated subsequences. In this case,

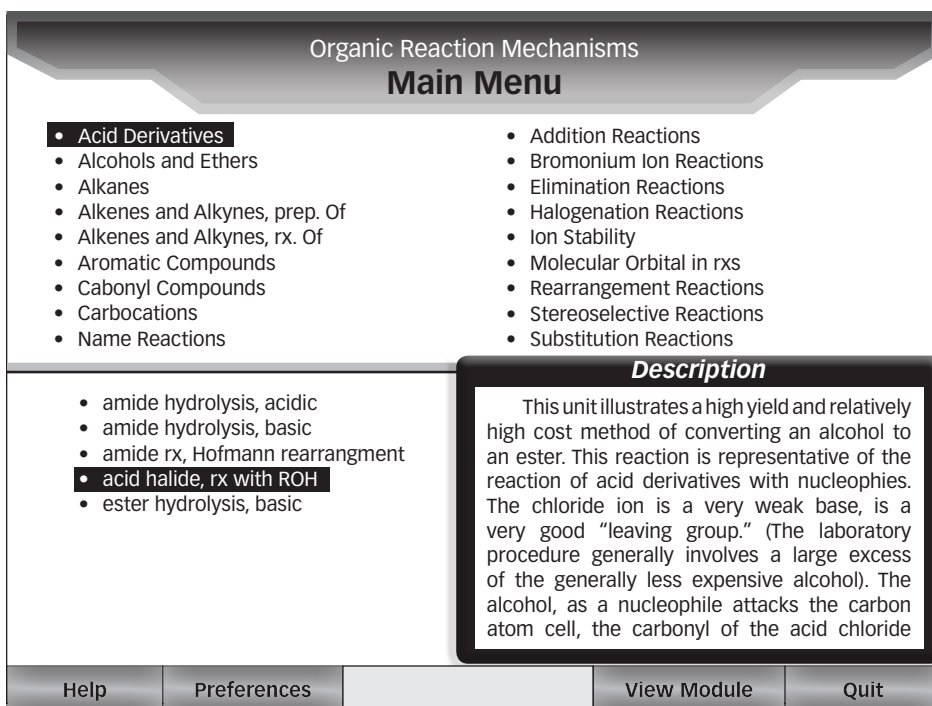


FIGURE 10. Main menu (acid derivatives: acid halide, rx with ROH)

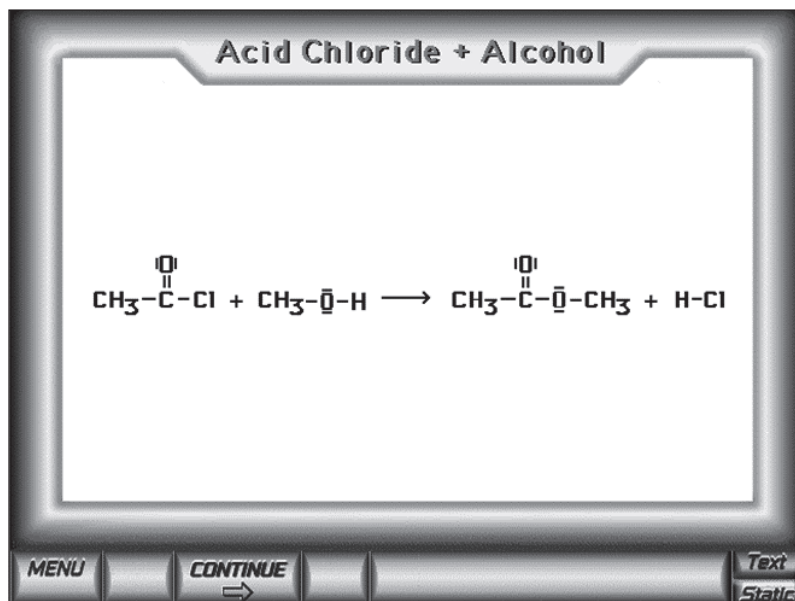


FIGURE 11. ORM Acid Chloride + Alcohol animation module

understanding the file naming pattern—that the file name for each DXR file used abbreviations of chemical compounds that then corresponded to a specific educational tutorial—helped the researcher better understand how each DXR file related to the larger ORM program.

Documenting a Format

To preserve knowledge learned during the research process about the dependency shared between Director DXR-protected movies and Windows projectors, the researcher created two entries in the Medusa Digital Content Format Registry: one for Director-protected movies (DXR) and another for Director Windows projectors (EXE). The researcher then supplemented these with a third entry to provide broader documentation of the Director format as a whole. The descriptive content names assigned to each entry were as follows:

Macromedia / Adobe Director (Shockwave) Protected Movie⁴⁵

Macromedia / Adobe Director (Shockwave) Windows Projector⁴⁶

Macromedia / Adobe Director (Shockwave) Format⁴⁷

For each entry, the researcher recorded PRONOM IDs, native file extensions, and related formats, as well as a rendering profile, administrative notes, and a recommended preservation and curation policy.

As an example, Figure 12 displays the core fields and a compact view of the file-rendering profiles for the Macromedia / Adobe Director (Shockwave) Protected Movie registry entry. The two most appropriate PRONOM IDs—one describing Director for PC files and the other describing Director for Macintosh files—are linked to the entry. The policy summary consists of an abbreviated description of the DXR-protected movie file type, a list of known dependencies, a statement of confidence in preservation strategies such as migration and normalization, and a recommended curatorial practice. The DXR extension was recorded as the protected movie's only native file extension, and the Director Windows Projector and Director Format registry entries were linked to the Director-protected movie entry as related formats.

Digital Content Format:

Macromedia / Adobe Director (Shockwave) Protected Movie

Name: Macromedia / Adobe Director (Shockwave) Protected Movie

Pronom: fmt/341 (Director - Macintosh), x-fmt/341 (Director - PC)

Policy Summary: Director protected movies (.DXR) are compressed and encrypted multimedia/ animation files. They have a dependency on Windows or Macintosh run-time environments known as “projectors,” which often take the form of a software program or application distributed onto external media such as a floppy disc, CD-ROM, DVD, or Apple mobile device. Due to the compressed and encrypted nature of Director protected movies, their technical dependency on Director projectors, and the complexity of the multimedia and animated content in which they may encode, there are currently no trusted migration or normalization solutions for Director .DXR files. [Emulation of] A previous or obsolete version of a Windows or Mac OS may be necessary to successfully render and play Director projectors and their linked .DXR protected movies. When curating Director protected movies for Preservation, Nearline, or Online access it is recommended practice to preserve the entire contents and original arrangement of the Director projector—i.e., the software program or application originally on floppy disc, CD-ROM, or DVD—to which the Director .DXR files belong, including necessary program files and any linked multimedia files (e.g., image, audio, codec, A/V formats).

Logical Extensions: dxr

Related Digital Content Formats: Macromedia / Adobe Director (Shockwave) Format , Macromedia / Adobe Director (Shockwave) Windows Projector

Rendering Profiles:			
Show <input checked="" type="checkbox"/> entities		Filter results:	
Name	Status	Software (Version)	OS (Version)
Macromedia / Adobe Director Windows Projector on (Windows OS dependent on .exe file version)	active	()	Microsoft Windows (OS version dependent on .EXE file version)
Organic Reaction Mechanisms, School Edition 2.0 CD-ROM on Microsoft Windows 3.1	active	Organic Reaction Mechanisms, School Edition (2.0)	Microsoft Windows (3.1)
application/x-dosexec	active	Schweitzer Engineering Laboratories (SEL) Manufacturing Virtual Tour (Version 2, June 2001)	Status

Showing 1 to 3 of 3 entries

FIGURE 12. Macromedia/Director (Shockwave) Protected Movie: core fields and rendering profiles

Rendering Profile:

Organic Reaction Mechanisms, School Edition 2.0 CD-ROM on Microsoft Windows 3.1

Index

Status: active

Software: Organic Reaction Mechanisms, School Edition

Software Version: 2.0

Os Environment: Microsoft Windows

Os Version: 3.1

Notes: Organic Reaction Mechanisms, School Edition (ORMS) contains interactive educational modules of animated organic chemical reactions and was authored by Andrew F Montana and Jeffery R. Buell of California State University, Fullerton. It was published on CD-ROM for Windows 3.1 circa 1995 by Falcon Software, Inc. ORMS's projector, ORMSFULL.EXE, is a 16-bit New Executable (NE). New Executables are designed for 16-bit platforms—e.g., multitasking MS-DOS 4.0, Windows 1.0 - Windows 3x. They are also compatible with the 32-bit Windows 9x platforms: Windows 95, Windows 98, and Windows ME. To play the ORMS CD-ROM, transfer a copy of the entire ORM directory from the Stanley Smith Papers (https://medusa.library.illinois.edu/cfs_directories/336861) to the target Windows 3.1, or compatible OS, file system. Launch Organic Reaction Mechanisms by double-clicking on the executable file ORMSFULL.EXE.

Digital Content format: New Executable

Digital Content format: Macromedia / Adobe Director (Shockwave) Format

Digital Content format: Macromedia / Adobe Director (Shockwave) Windows Projector

Digital Content format: Macromedia / Adobe Director (Shockwave) Protected Movie

Content types:
• application/x-director
• application/x-dosexec

File Extensions:
• dll
• dxr
• exe

FIGURE 13. "Organic Reaction Mechanisms, School Edition 2.0 CD-ROM on Microsoft Windows 3.1" rendering profile

The researcher created two initial rendering profiles: 1) a profile for rendering Windows projectors in general⁴⁸ and 2) a profile for specifically rendering the Organic Reaction Mechanisms, School Edition (ORMS) CD-ROM.⁴⁹ A follow-up investigation of a DXR file located in another collection yielded an additional profile for the Schweitzer Engineering Laboratories (SEL) Manufacturing Virtual Tour CD-ROM.⁵⁰ The rendering profile for the ORMS CD-ROM shown in Figure 13 consists of ORMS software and Windows 3.1 versioning information, a notes field containing background information about ORMS and detailing Windows OS compatibility and rendering instructions, and a list of the specific extensions and content types of the files required to run the software. In this case, the files required to run the ORMS CD-ROM include Director-protected movie (DXR) files, a Director Windows projector (New Executable EXE) file, as well as Windows Dynamic-Link Library (DLL) files. The file format field lists entries within the Medusa Digital Content Format Registry in which the rendering profile is associated. Thus, in addition to being applicable to the Protected Movie registry entry, the ORMS CD-ROM rendering profile also applies to and appears on the New

Notes

Add

kgermeck - 2019-03-11

Edit

Delete

[Description & Rendering]

Macromedia/Adobe Director is a multimedia authoring and internet publishing tool for Shockwave, the web-based application and video game platform (not to be confused with the similar web-based platform, Adobe Shockwave Flash). For a more detailed description of Director, please refer to the entry for the Macromedia / Adobe Director (Shockwave) File Format: https://medusa.library.illinois.edu/file_formats/215

A Protected Movie, .DXR, is a Director movie that has been compressed and encrypted in order to protect the intellectual property—the code and scripts—of a movie from being accessible once it has been distributed for publication. DXR files cannot be rendered or opened by the Director software program, nor can they be properly played within a web browser using the Shockwave plug-in or Shockwave Player (see resource #5 in the Resources section).

DXR files are only playable in association with a Director Projector (a run-time environment, typically in the form of an executable Windows software program or Mac application) that has been linked to a .DXR file or a group of .DXR files. (See the Macromedia / Adobe Director (Shockwave) Windows Projector registry entry: https://medusa.library.illinois.edu/file_formats/217). The Director projector associated with the .DXR file(s) must be launched first. Once within the program, a menu selection or interactive activity (i.e., a .DXR file) can be selected and played. For more information about rendering .DXR files see the Macromedia / Director Windows Projector rendering profile: https://medusa.library.illinois.edu/file_format_profiles/189.

kgermeck - 2019-03-11

Edit

Delete

[Director Protected Movies (.DXR) in Medusa]

There are currently forty-seven unique .DXR protected movie files within the Stanley Smith Papers and one .DXR protected movie file in the College of Engineering Corporation Files.

1) The .DXR files in the Smith Papers are linked to the Director projector ORMSFULL.EXE, which is the executable file for the Organic Reaction Mechanisms, School Edition (ORMS) CD-ROM.

The Director projector for the ORMS CD-ROM, ORMSFULL.EXE, is located within the ORM directory: https://medusa.library.illinois.edu/cfs_directories/328301. The .DXR files for the ORMS CD-ROM are located within the ORM subdirectory DATA: https://medusa.library.illinois.edu/cfs_directories/328302. The ORMS .DXR files contain animation sequences that illustrate organic chemical reactions. For rendering instructions, refer to the "Organic Reaction Mechanisms, School Edition 2.0 CD-ROM on Microsoft Windows 3.1" rendering profile.

Stanley Smith incorporated a number of individual animated sequences from ORMS as .GIF images into his web-based course, Organic Chemistry, published by Falcon Software in 2001.

A sample .GIF image from Smith's course containing animation from ORMS is viewable here: <https://archives.library.illinois.edu/erec/University%20Archives/1505050/Organic/Mechanism/RX/SN2Animation.htm>.

2) The one .DXR file in the College of Engineering Corporation Files is linked to the Director projector Projector.exe, which is the executable file for the Schweitzer Engineering Laboratories (SEL) Manufacturing Virtual Tour CD-ROM. (The CD-ROM is also linked to three .mpg and one .mpeg video files).

The SEL CD-ROM's projector (Projector.exe) and its one .DXR file (detect.dxr) are located in the same directory: https://medusa.library.illinois.edu/cfs_directories/1311204. Detect.DXR consists of a screen resolution settings switcher, which runs upon launching the SEL CD-ROM. Once a screen resolution selection is made, the .mpg/.mpeg video that corresponds to the selected screen resolution setting is launched and played. For rendering instructions, refer to the "Schweitzer Engineering Laboratories (SEL) Manufacturing Virtual Tour CD-ROM on Microsoft Windows 10" rendering profile.

NOTE: For more detailed information about the ORMS and SEL CD-ROM projectors, refer to the entry for the "Macromedia / Adobe Director (Shockwave) Windows Projector": https://medusa.library.illinois.edu/file_formats/217

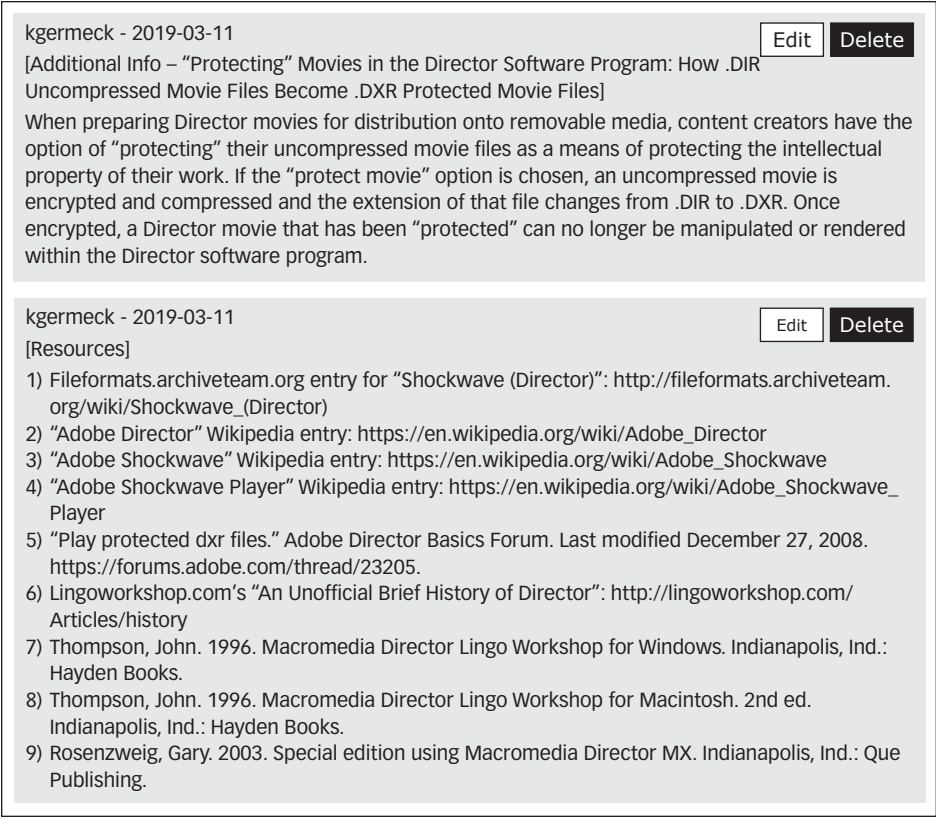


FIGURE 14. Macromedia / Adobe Director (Shockwave) Protected Movie notes files

Executable, Macromedia / Adobe Director (Shockwave) Windows Projector, and Macromedia / Adobe Director (Shockwave) Format registry entries.

To accommodate and shape a contextually rich narrative about the local qualities and behavior of a format, the authors developed a consistent series of designated sections, or tags, for a registry entry’s administrative notes. For example, the notes for the Protected Movie entry, shown in Figure 14, are organized into the four following sections: 1) Description and Rendering, 2) Director Protected Movies in Medusa, 3) Additional Info, and 4) Resources.

The “Description and Rendering” section provides a summary of the Director software program and the DXR-protected movie file type and explains DXR’s dependence on Director projectors. The “Director Protected Movies in Medusa” section highlights the Medusa collections and directory locations in which Director DXR files and their linked projectors appear and also provides descriptions of the content encoded within the DXR files. The “Additional Info” section addresses the significance of the Director software program’s “Protect Movie” feature, explaining how that process results in the conversion of an uncompressed movie (DCR) file into a protected movie (DXR) file. The “Resources”

Normalization Paths

Add

Name	Input Extension	Output Extension	Output Format
------	-----------------	------------------	---------------

Attachments

Add

Filename	Description
SampleDXR_text_and_hex_editor_views.pdf	Demonstrates how to retrieve and confirm opening bit sequence and file version info
ORM_ACCRL_ROH_AnimationModule.mp4	Video capture illustrating the animated content of the ACCRL_ROH.DXR: Acid Chloride + Alcohol Animation Module in the Organic Reaction Mechanisms, School Edition (ORMS) CD-ROM.

FIGURE 15. Macromedia / Adobe Director (Shockwave) Protected Movie normalization paths and attachments

section lists citations and/or hyperlinks to the print and online sources most relevant to researching and documenting the local properties and needs of Director-protected movie files stored within Medusa. Due to the compressed and encrypted nature of Director-protected movies, their technical dependency on Director projectors, and the complexity of the multimedia and animated content in which they may encode, no suitable normalization path was defined for the format.

When appropriate, documentation created during the process of researching a format (e.g., screenshots, video captures, typed instructions or notes) may be attached to a registry entry. Attachments to the Protected Movie entry, as shown in Figure 15, include screenshots of identifying the bit-level header and file version of a sample DXR file and a video capture illustrating the animated content of a DXR file from the OMRS CD-ROM.

The policy summary for the Protected Movie format, highlighted in Figure 16, emphasizes the technical complications (e.g., proprietary encryption, file dependency) preventing confidence in the migration or normalization of DXR files and states that emulation may be necessary to successfully render protected movies and their corresponding projectors. Furthermore, it addresses the arrangement and access of DXR files and recommends maintaining the entire contents and original file directory structure of a software program or application associated with a Director projector and DXR files. A preservation or access copy of a set of DXR files, that is, should include the Director projector, necessary program files, and any externally linked multimedia files on which the program or application may depend.

The DXR files researched and documented here provide a rich example of the interdependent, and often obfuscated, relationships that may occur

Downloaded from https://prime-pdf-watemark.prime-prod.pubfactory.com/ at 2025-07-01 via free access

Name: Macromedia / Adobe Director (Shockwave) Protected Movie

Pronom: fmt/341 (Director - Macintosh), x-fmt/341 (Director - PC)

Policy Summary: Director protected movies (.DXR) are compressed and encrypted multimedia/animation files. They have a dependency on Windows or Macintosh run-time environments known as “projectors,” which often take the form of a software program or application distributed onto external media such as a floppy disc, CD-ROM, DVD, or Apple mobile device. Due to the compressed and encrypted nature of Director protected movies, their technical dependency on Director projectors, and the complexity of the multimedia and animated content in which they may encode, there are currently no trusted migration or normalization solutions for Director .DXR files. [Emulation of] A previous or obsolete version of a Windows or Mac OS may be necessary to successfully render and play Director projectors and their linked .DXR protected movies. When curating Director protected movies for Preservation, Nearline, or Online access it is recommended practice to preserve the entire contents and original arrangement of the Director projector—i.e., the software program or application originally on floppy disc, CD-ROM, or DVD—to which the Director .DXR files belong, including necessary program files and any linked multimedia files (e.g., image, audio, codec, A/V formats).

Logical Extensions: dxr

Related Digital Content Formatst: Macromedia / Adobe Director (Shockwave) Format ,
Macromedia / Adobe Director (Shockwave) Windows Projector

FIGURE 16. Macromedia / Adobe Director (Shockwave) Protected Movie policy summary, with preservation and curation statements highlighted

between different file types and formats within a complex digital object such as a CD-ROM application from the 1990s. The research process reinforced for the authors the importance of maintaining the integrity of contextual file system relationships for access and preservation.

Analysis

The example of Macromedia files from the early 1990s is instructive for several reasons. This particular software and its several file formats were popular during the years in which they dominated their market. Twenty-five years after the fact, however, they are so rare as to demand a research effort to understand how they worked, an effort that eventually required the emulation of a bygone operating system to bring certain files back to life. John Henry Thompson, the largely unsung luminary of computer programming who created this software suite, has expressed concern in recent years about the future accessibility of Macromedia files and software, saying that content created “not just with Director, but with Flash and a lot of other digital content—is just going to slowly evaporate from our heritage because it is not available anymore.”⁵¹ Notably, Thompson maintains a website dedicated to efforts to keep obsolescent Director software alive for the cultural heritage community⁵² and has openly courted digital archivists as allies to his cause.

The Macromedia example suggests that research into file formats, legacy software, and proficiency in software emulation and other skills related to so-called digital archaeology will become increasingly necessary for archivists and their colleagues in digital preservation services during the appraisal process. That is, the ability to bring historical software and operating systems to life is often required to even assess what certain digital content is, especially for the archivist who would like to perform a hands-on appraisal of born-digital records dating from twenty years ago or more. Director files are just one family of file formats among many, and the authors could have just as easily selected any among hundreds of others that exist in their holdings.

To what extent is a general international file format registry sufficient for the digital preservation knowledge needs of a single organization, beyond the essential functions of cataloging format signatures and other identifying traits? As Trevor Owens has written, digital preservation is not primarily a technical problem to be solved with adequate software; rather, preservation encompasses “myriad local problems contingent on what different communities value,” and, as such, it must be conducted differently depending on its context.⁵³ This means that questions of preservation and access must be driven by the needs of an archives’ constituents. A local format registry thus affords curators a unique opportunity to understand their digital materials and better address the access needs of their user communities. It also helps to close the gap between curators and users who seek to discover, access, and use special collections and the files that compose them.

This study suggests that, especially for archivists managing historical collections of heterogeneous born-digital materials, maintaining a knowledge base of local digital content types has value because a file format is not a single, bounded thing; it is an approximate notion that encompasses a multiplicity of different types of information containers. And, as the DXR file example illustrates, context is critical, especially in informing appraisal and curatorial decisions. An international file format registry may provide reliable reference information on what makes a PDF a PDF, but it may not provide information about why a particular type of locally held PDF is difficult to open without specialized software. By complementing the maintenance of universal format registries with local knowledge stores, preservation managers may identify, define, and manage the content types that occur most frequently in their own collections.

While commonalities between born-digital materials can be found across repositories and between institutions, archivists ought to always anticipate local idiosyncrasies. For this reason, we suggest a layered model of file format information (see Figure 17), relying on reference services when possible to characterize file formats in general and contributing knowledge back to these when possible, while cultivating local records to allow curators, preservation

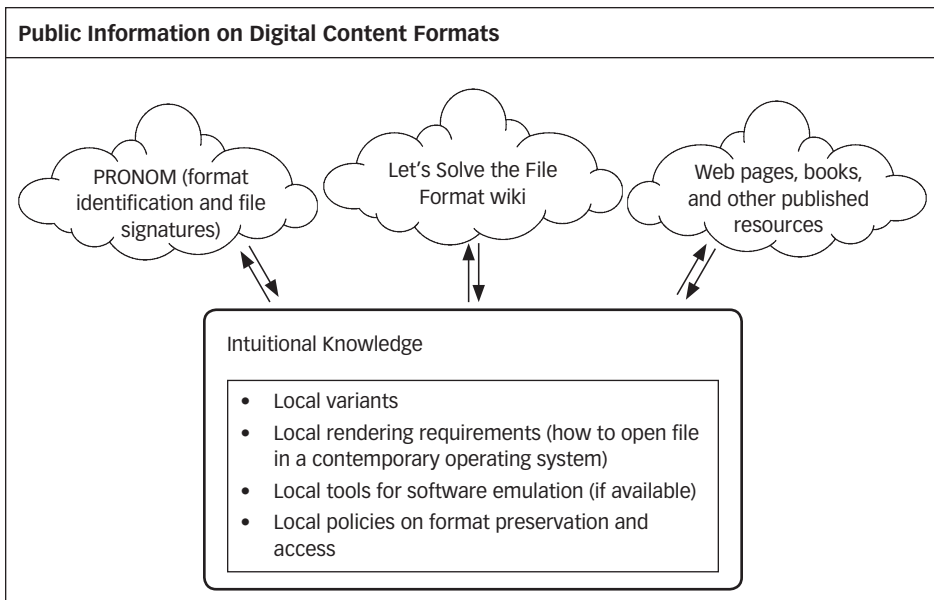


FIGURE 17. Integrating publicly available information on digital content formats with institutional knowledge and policies

managers, and patrons to better understand how to access local strains that recur in collections of interest.

Much could be gained from sharing local profiles with peer institutions, for example, in a consortial setting, especially to leverage staff expertise across organizations with similar areas of collecting focus. The question is still open regarding just how local these ancillary registries ought to be, or how they might best be shared. Regionally, the Big Ten Academic Association's Digital Preservation Group has discussed sharing knowledge in file formats across the group, but, to date, no progress has been made in implementing a shared service.

Initiating this model at the organizational level comes with a cost. The learning curve for doing research into digital formats is steep for staff who are not already familiar with the brief but daunting history of bygone personal computer technologies. Once trained, staff must dedicate time on an ongoing basis to monitoring digital content formats as these are acquired by the repository and must then communicate their findings to their colleagues to ensure some consistency of curatorial practice. Nor is it feasible for every archives in the world to build independent local digital format registries. Looking ahead, experts in the field of archives and digital preservation ought to consider how best to integrate these tiers of general and local knowledge about digital content types and build them into software available for digital curation, as well as workflows for appraisal, arrangement, and description. Data-sharing and aggregation models from other areas of archival practices, such as the Social

Networks and Archival Context project,⁵⁴ may be worthy of consideration, particularly if they can be built on top of or as an extension to an existing service, such as PRONOM.

For the time being, the authors maintain that the practices outlined in this article have a place in archives, even if their research-intensive nature contrasts with the type of time-saving affordances most people hope to gain from digital formats. Given the state of digital preservation practice, archivists cannot escape the hands-on exigencies of digital preservation work when managing born-digital collections. While certain tools exist to aid in the automation of individual tasks related to appraisal, arrangement, and description, the field is far from possessing a device that automatically scans a directory of files and identifies precisely their format (and what variety of that format), how to access their content, what other files relate to them (in a technical as well as in a documentary context),⁵⁵ which files contain sensitive personal information, and so on. In the future, tools will likely exist to further simplify archival workflows, but the present state of technology and descriptive practice still requires a significant amount of bitstream-level appraisal to truly understand the nature of digital content acquired. And this appraisal depends on the context of the materials, which will always be unique for different *fonds*, records series, or manuscript collections.

As appraisal and processing practices evolve in light of the data afforded by our Digital Content Format Registry, we anticipate the registry will evolve to be better attuned and integrated into local workflows. Appraisal requires consideration of a variety of factors, and no two file formats will necessarily be appraised the same way based on their context. While it is true that knowledge about file dependencies and relationships will considerably enhance archivists' understanding of how the files originally operated and, in some cases, interacted with each other, the degree to which this information will influence appraisal and curatorial decisions will inevitably vary depending on the context of the materials. Furthermore, while digital formats lend themselves best to streamlined bulk automation and streamlined processes when they have been produced under controlled circumstances leading to consistent, uniform results, born-digital archives are more likely than traditional collections of digitized library materials to contain either peculiar file types or common file types produced in an inconsistent, idiosyncratic manner.⁵⁶ For this reason, workflows based upon retaining such files in their original order and format will do a suitable job of respecting archival values.

Looking to the future, we intend to use our Digital Content Format Registry to inform research into the use of emulation for content access.⁵⁷ Particularly when curating collections of materials from "production" environments related to audio and video and emergent formats such as three-dimensional images and

virtual reality, the exported final content rarely adequately represents the significant properties associated with digital production workstation files. These files and their respective software-rendering environments are complex and require significant knowledge to access them in their original form, making them strong candidates for emulation as an access solution. However, scaling emulation services for institutional digital preservation and curation practice will involve significant knowledge development and sharing. In this context, local digital content format registries will serve as valuable building blocks for knowledge gathering and retention. It is even possible that research outlined in this article could help open the door to a common data-sharing model for digital content formats and the eventual sharing of locally generated format data among archivists and preservationists from multiple organizations.

Conclusions and Next Steps

We have demonstrated a working model for an institution-level digital content format registry. We found that it is feasible to implement a preservation strategy around cultivating local information on specialized digital content types. The collection of this knowledge requires human intervention and, at this time, a meaningful amount of labor. If, one day, all can be done algorithmically, such a service would by its very nature depend on thoughtful human work to extend the range of knowledge beyond that supplied by format characterization tools. In this sense, a strategy focused on developing local knowledge regarding file access, rendering, and preservation pathways is an essential element of the work that all digital archivists undertake. Nor is it specific to a digital preservation approach that favors file format normalization over emulation. Rather, a local format registry is of pragmatic value whether or not institutions adopt emulation, because local registries inform decision-making, and do not prescribe outcomes so much as help ensure that outcomes align with the needs of a designated archives and its users.

At the time of publication, the authors' Digital Content Format Registry featured 250 entries, most of which were added by current and past staff or graduate student workers in the library's Preservation Services unit. The role of the registry in local archival workflows will undoubtedly evolve as archivists integrate it into appraisal and curatorial frameworks. The unique, local context in which the born-digital materials at University of Illinois at Urbana-Champaign are created must be taken into account along with the growing institutional knowledge about the variety of the file format's characteristics that comprise those materials. We anticipate our registry will lead to more holistic appraisal decisions augmented by enhanced knowledge about file formats. These informed appraisal decisions will lead to curatorial decisions that

better enable users to more authentically experience and interact with born-digital materials. While their own implementation is young, the authors see many benefits to an organizational file format registry. As such, we encourage archivists and digital preservationists to consider how best to meaningfully integrate the information contained in international file format registries with local knowledge gained from the hands-on curation of born-digital materials.

NOTES

- ¹ Donald Waters and John Garrett, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, the Commission on Preservation and Access and the Research Libraries Group (Washington, DC: Council on Library and Information Resources, May 1, 1996), <https://clir.wordpress.org/wp-content/uploads/sites/6/pub63watersgarrett.pdf>, captured at <https://perma.cc/EU3P-GAER>.
- ² Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Washington, DC: Council on Library and Information Resources, January 1999). Rothenberg points out that “Digital documents are inherently software-dependent” (p. 8), and thus argues that the preservation of software is not only of great importance to the preservation of digital documents, but that software emulation is “the only approach yet suggested to offer a true solution to the problem of digital preservation” (p. v).
- ³ A representative example is David Bearman, “Reality and Chimeras in the Preservation of Electronic Records,” *D-Lib Magazine* 5, no. 4 (1999), <http://www.dlib.org/dlib/april99/bearman/04bearman.html>. For further discussion of the emulation debate, see also Stewart Granger, “Emulation as a Digital Preservation Strategy,” *D-Lib Magazine* 6, no. 10 (2000), <https://doi.org/10.1045/october2000-granger>.
- ⁴ For a thoughtful investigation of remaining barriers and opportunities for emulation in digital preservation, see David S. H. Rosenthal, “Emulation and Virtualization as Preservation Strategies,” Andrew W. Mellon Foundation Research Reports, October 2015, https://mellon.org/media/filer_public/0c/3e/0c3eee7d-4166-4ba6-a767-6b42e6a1c2a7/rosenthal-emulation-2015.pdf, captured at <https://perma.cc/B5ZH-DG8A>. See also “Fostering a Community of Practice: Software Preservation in Libraries and Archives,” Software Preservation Network, <http://www.softwarepreservationnetwork.org/fcop>; and, for an excellent look at emulation in practice, the Internet Archive’s “Handheld History,” featuring web-accessible emulated versions of handheld games and computing devices, <https://archive.org/details/handheldhistory>.
- ⁵ J. Rog and C. Van Wijk, “Evaluating File Formats for Longterm Preservation,” *Koninklijke Bibliotheek* 2 (2008): 12–14.
- ⁶ David Pearson and Colin Webb, “Defining File Format Obsolescence: A Risky Journey,” *International Journal of Digital Curation* 3 (2008): 89–106.
- ⁷ Richardson Anderson, Hannah Frost, Nancy Hoebelheinrich, and Keith Johnson, “The AIHT at Stanford University: Automated Preservation Assessment of Heterogeneous Digital Collections,” *D-Lib Magazine* 11, no. 12 (2005): 10, <https://doi.org/10.1045/december2005-johnson>.
- ⁸ Andreas Stanescu, “Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology,” *OCLC Systems & Services* 21, no. 1 (2005): 61–81.
- ⁹ Kyle Rimkus, Thomas Padilla, Tracy Popp, and Greer Martin, “Digital Preservation File Format Policies of ARL Member Libraries: An Analysis,” *D-Lib Magazine* 20, nos. 3–4 (2014), <https://doi.org/10.1045/march2014-rimkus>.
- ¹⁰ For an in-depth exploration of this theme, see Trevor Owens, *The Theory and Craft of Digital Preservation* (Baltimore: Johns Hopkins University Press, 2018), in particular chapters 2 and 3.
- ¹¹ Kevin DeVorsey and Peter McKinney, “Digital Preservation in Capable Hands: Taking Control of Risk Assessment at the National Library of New Zealand,” *Information Standards Quarterly* 22, no. 2 (2010): 41–44.

- ¹² InterPARES 1 Preservation Task Force, "Preservation Task Force Report" (2002), http://www.interpares.org/book/interpares_book_f_part3.pdf. Emphasis added.
- ¹³ Christoph Becker, "Metaphors We Work By: Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems," *Archivaria* 85 (Spring 2018): 6–36, <https://archivaria.ca/index.php/archivaria/article/view/13628>.
- ¹⁴ Peter McKinney et al., "Reimagining the Format Model: Introducing the Work of the NSLA Digital Preservation Technical Registry," *New Review of Information Networking* 19, no. 2 (2014): 96–123, <https://doi.org/10.1080/13614576.2014.972718>.
- ¹⁵ Other partners in the Preservation Actions Registries include Artefactual, Arkivum, Preservicea, and Jisc. For more information, see Par, <http://parcore.org>.
- ¹⁶ Educopia Institute, "Scaling Emulation as a Service Infrastructure (EaaS) (subcontract), <https://educopia.org/emulation-as-a-service-eaasi>.
- ¹⁷ Gareth Kay, Libor Coufal, and Mark Pearson, "Backing Up Digital Preservation Practice with Empirical Research: The National Library of Australia's Digital Preservation Knowledge Base," *Alexandria: The Journal of National and International Library and Information Issues* 27, no. 2 (2017): 66–82, <https://doi.org/10.1177/0955749017724630>.
- ¹⁸ Christopher J. Prom and Ellen D. Swain, "From the College Democrats to the Falling Illini: Identifying, Appraising, and Capturing Student Organization Websites," *American Archivist* 70, no. 2 (2007): 344–63, <https://doi.org/10.17723/aarc.70.2.c8121767x9075210>.
- ¹⁹ Christopher J. Prom, "Sabbatical Report," *Practical E-Records* (blog), July 14, 2010, <http://e-records.chrisprom.com/reports/sabbatical-report-illinois>, captured at <https://perma.cc/YJG5-DDJG>.
- ²⁰ Much of this testing was chronicled through Chris Prom's *Practical E-Records* blog at <http://e-records.chrisprom.com>.
- ²¹ Some content is publicly available at <https://medusa.library.illinois.edu>, although access to the system itself is password restricted.
- ²² The Medusa system's source code is available at <https://github.com/Medusa-project>. While Medusa software is not supported in such a way as to be easily adopted by other institutions, this is beginning to change, as its developers recently refactored Medusa for deployment in the Amazon Web Services cloud.
- ²³ An earlier publication (Kyle Rimkus and Thomas G. Habing, "Medusa at the University of Illinois at Urbana-Champaign: A Digital Preservation Service Based on PREMIS," in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, 49–52 (Indianapolis: ACM/IEEE-CS, 2013), states that Medusa was built using the Fedora digital asset management architecture; while this was the plan at the time, it is no longer the case.
- ²⁴ Consultative Committee for Space Data Systems, "Reference Model for an Open Archival Information System (OAIS)," CCSDS Secretariat, 2012, <https://public.ccsds.org/pubs/650x0m2.pdf>, captured at <https://perma.cc/M86F-QS7A>. Originally conceived in 2002, the OAIS model was updated in 2012 and subsequently codified as ISO standard 14721, <https://www.iso.org/standard/57284.html>.
- ²⁵ Medusa takes cues from archival management systems such as Archon, the Archivist's Toolkit, ICA-AtoM, and ArchivesSpace by providing curators with a web-accessible interface for managing preservation actions over time. It features forms for editing descriptive, administrative, and rights metadata associated with collections; downloading files or batches of files to local file servers; tracking of preservation events, file provenance, and file statistics; on-demand verification of file fixity (md5 checksum values) for files or batches of files and ongoing checksum verification of all files every ninety days; on-demand extraction of technical metadata (using the File Information Tool Set, or FITS) for files or batches of files; and, as we describe further in this article, a Digital Content Format Registry with tools for documenting local file format knowledge and policies.
- ²⁶ The local Digital Content Format Registry referenced in this article is openly available at University Library, University of Illinois at Champaign-Urbana, Medusa, https://medusa.library.illinois.edu/file_formats.

- ²⁷ Some precedent exists for this decision; for example, the US Library of Congress uses similar terminology in its excellent “Sustainability of Digital Formats” resource, <https://www.loc.gov/preservation/digital/formats>.
- ²⁸ See “Pronom,” the UK National Archives, <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=686>.
- ²⁹ Kyle R. Rimkus and Scott D. Witmer, “Identifying Barriers to File Rendering in Bit-Level Preservation Repositories: A Preliminary Approach,” in *Proceedings of the 13th International Conference on Digital Preservation* (Bern, Switzerland: Swiss National Library, 2016), 121–28, <https://www.ideals.illinois.edu/handle/2142/91660>.
- ³⁰ University Library, University of Illinois at Champaign-Urbana, Medusa, https://medusa.library.illinois.edu/file_formats.
- ³¹ Documentation related to the Stanley Smith Papers (Born-digital Records, Digital Surrogates and Audiovisuals), 1989–2011, 2014, located in Stanley Smith Papers, 1957–2006, 15/5/50, University of Illinois Holdings Database, <https://archives.library.illinois.edu/archon/?p=digitallibrary/digitalcontent&id=6381>.
- ³² Director was originally developed under the name “VideoWorks” by MacroMind in 1985 and further developed by Macromedia during the late 1980s through the early 2000s. Adobe managed the development and production of Director from 2005 to 2017. See also “Shockwave (Director),” Format entry, Let’s Solve the File Format Problem, last modified March 5, 2019, [http://fileformats.archiveteam.org/wiki/Shockwave_\(Director\)](http://fileformats.archiveteam.org/wiki/Shockwave_(Director)).
- ³³ The Creative Cloud Team, “The Future of Adobe Contribute, Director and Shockwave,” *Adobe Blog* (blog), January 27, 2017, <https://theblog.adobe.com/the-future-of-adobe-contribute-director-and-shockwave>.
- ³⁴ Notepad++, <https://notepad-plus-plus.org>.
- ³⁵ “Macromedia Director PC,” Format entry, x-fmt/341, the Technical Registry PRONOM, United Kingdom National Archives, last modified January 18, 2011, <https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=505>. The opening bit sequence for a Macromedia Director for Windows file is 52494658 (RIFX).
- ³⁶ John Thompson, *Macromedia Director Lingo Workshop for Macintosh*, 2nd ed. (Indianapolis: Hayden Books, 1996), 278, 287, 293.
- ³⁷ Thompson, *Macromedia*, 284, 302.
- ³⁸ Thompson, *Macromedia*, 278, 282–83.
- ³⁹ James Newton, “Playing Protected DXR Files,” in response to user “mgbyrnes,” Adobe Community Director Basics Forum, last modified December 27, 2008, <https://forums.adobe.com/message/1006680#1006680> [link inactive]. According to this response on an Adobe Community Director Basics Forum thread from 2008, if a DXR file is written with scripts made available by the Xtras feature in Director—which, the responder notes, is incompatible with Shockwave—the file will cause errors within Shockwave Player and will not play.
- ⁴⁰ Understanding records in their documentary context—i.e., the bond that exists between records belonging to the same *fonds*—is important to the records’ meaning. See Luciana Duranti, “The Archival Bond,” *Archives and Museum Informatics* 11 (1997): 213–18.
- ⁴¹ “Windows New Executable,” Format entry, x-fmt/410, the Technical Registry PRONOM, United Kingdom National Archives, last modified February 17, 2006, <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=775&strPageToDisplay=signatures>. The opening bit sequence of a New Executable is 4d5a (MZ); the offset header is 4e45 (NE).
- ⁴² In 1994, when Macromedia Director 4.0 was released, the newest Microsoft 16-bit platform was Windows 3.1.
- ⁴³ Andrew F. Montana and Jeffrey R. Buell, *Organic Reaction Mechanisms*, version School Edition 2.0 (Wellesley, MA: Falcon Software, 1995). *Organic Reaction Mechanisms*, School Edition (ORMS) contains interactive educational modules of animated organic chemical reactions.
- ⁴⁴ Stanley Smith incorporated a number of individual animated sequences from ORMS as GIF images into his web-based course in organic chemistry, published by Falcon Software in 2001. A sample GIF image from Smith’s course containing animation from ORMS is viewable at <https://archives>

- .library.illinois.edu/erec/University%20Archives/1505050/Organic/Mechanism/RX/SN2Animation.htm.
- ⁴⁵ “Macromedia / Adobe Director (Shockwave) Protected Movie,” Format entry, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_formats/216.
- ⁴⁶ “Macromedia / Adobe Director (Shockwave) Windows Projector,” Format entry, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_formats/217.
- ⁴⁷ “Macromedia / Adobe Director (Shockwave) Format,” Format entry, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_formats/215.
- ⁴⁸ “Macromedia / Adobe Director Windows Projector on (Windows OS Dependent on .Exe File Version),” Rendering profile, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_format_profiles/189.
- ⁴⁹ “Organic Reaction Mechanisms, School Edition 2.0 CD-ROM on Microsoft Windows 3.1,” Rendering profile, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_format_profiles/188.
- ⁵⁰ “Schweitzer Engineering Laboratories (SEL) Manufacturing Virtual Tour CD-ROM on Microsoft Windows 10,” Rendering profile, Medusa Digital Content Format Registry, University Library, University of Illinois at Urbana-Champaign, https://medusa.library.illinois.edu/file_format_profiles/206.
- ⁵¹ Guest: John Henry Thompson, YouTube video, hosted by Daniel Shiffman, 31:26, March 27, 2018, posted by The Coding Train, <https://www.youtube.com/watch?v=DvS4h-1Eyu4>.
- ⁵² <http://www.johnhenrythompson.com/artifacts>.
- ⁵³ Owens, *The Theory and Craft of Digital Preservation*, 191.
- ⁵⁴ In this project, locally stored and managed data held in disparate formats are harvested, standardized, indexed, and shared via a cooperative service. See Daniel Pitti et al., “Social Networks and Archival Context: From Project to Cooperative Archival Program,” *Journal of Archival Organization* 12, nos. 1–2 (2015): 77–97, <https://doi.org/10.1080/15332748.2015.999544>.
- ⁵⁵ For some recent work on identifying and documenting dependencies for digital preservation, see Nikolaos Lagos, Marina Riga, Panagiotis Mitzias, Jean-Yves Vion-Dury, Efstratios Kontopoulos, Simon Waddington, Pip Laurenson, Georgios Meditskos, and Ioannis Kompatsiaris, “Dependency Modelling for Inconsistency Management in Digital Preservation—the PERICLES Approach,” *Information Systems Frontiers* 20 (2018): 7–19; and Nikolaos Lagos and Jean-Yves Vion-Dury, “Digital Preservation Based on Contextualized Dependencies,” in *DocEng ’16: Proceedings of the 2016 ACM Symposium on Document Engineering*, 35–44 (New York: Association for Computing Machinery, 2016).
- ⁵⁶ Rimkus and Witmer, “Identifying Barriers to File Rendering in Bit-Level Preservation Repositories: A Preliminary Approach,” 121–28.
- ⁵⁷ Co-author Tracy Popp is also participating in a project called Fostering a Community of Practice: Software Preservation and Emulation in Libraries, Archives and Museums, <https://www.softwarepreservationnetwork.org/fcop> and is focusing on the curation of born-digital files created in legacy music production and composition software category of content, which is notably challenging and idiosyncratic and not well represented in international registries.

ABOUT THE AUTHORS



Kyle Rimkus is preservation librarian with rank of associate professor at the University of Illinois at Urbana-Champaign Library, where he leads efforts to establish policies, technologies, and practices to ensure that digital collections will persist into the future. He has a master of science in library and information science and a master's in French literature from the University of Illinois at Urbana-Champaign, and a bachelor's in Germanic studies from the University of Illinois at Chicago.



Bethany Anderson is assistant professor and natural and applied sciences archivist in the University Archives at the University of Illinois at Urbana-Champaign. She serves as co-editor for the Archival Futures Series, which is copublished by the Society of American Archivists and the American Library Association, and as reviews editor for *American Archivist*. She holds an MSIS in archival studies and records management from the University of Texas at Austin and an MA in Near Eastern art and archaeology from the University of Chicago.



Karl E. Germeck served as resident librarian in digital preservation at the University of Illinois at Urbana-Champaign University Library from 2017 to 2020. He holds a master of science in library science with a concentration in archives and records management from the University of North Carolina at Chapel Hill (2016) and a master of science in American studies from Utah State University (2008). He is the recipient of the 2009 Lynn Interdisciplinary Graduate Fellowship in American Studies at Purdue University, where he conducted doctoral research on nineteenth- and twentieth-century public memorialization and the cultural memory of underrepresented populations in the United States.



Cameron C. Nielsen is a reference and instruction librarian at the Scranton and Wilkes-Barre campuses of Pennsylvania State University. He holds master's degrees in library and information science and in religion, both from the University of Illinois at Urbana-Champaign, where he worked as a graduate assistant in the University Library's Preservation Unit for two years. His research interests include topics in critical digital humanities, information literacy, and US religious history.



Christopher Prom is associate dean for digital strategies at the University of Illinois. He recently served as publications editor and chair of the Publications Board for the Society of American Archivists, and is a Fellow of the Society. He holds a PhD in history from the University of Illinois, with an emphasis in late Victorian social and labor history.



Tracy Popp is the digital preservation coordinator at the University of Illinois at Urbana-Champaign University Library. In support of preservation of and access to digital collections, Popp leads the library's born-digital collections preservation efforts. Popp holds a master of science of library and information science and a certificate of advanced study in library and information science from the University of Illinois at Urbana-Champaign.