

3D Data Repository Features, Best Practices, and Implications for Preservation Models: Findings from a National Forum

Juliet L. Hardesty, Jennifer Johnson, Jamie Wittenberg, Nathan Hall, Matt Cook, Zack Lischer-Katz, Zhiwu Xie, and Robert McDonald

This study identifies challenges and directions for 3D/VR repository standards and practices. As 3D technologies become more affordable and accessible, academic libraries need to implement workflows, standards, and practices that support the full lifecycle of 3D data. This study invited experts across several disciplines to analyze current national repository and preservation efforts. Outlined models provide frameworks to identify features, examine workflows, and determine implications of 3D data on current preservation models. Participants identified challenges for supporting 3D data, including intellectual property and fair use; providing repository system management beyond academic libraries; seeking guidance outside of academia for workflows to model.

Introduction

Three Dimensional (3D) and Virtual Reality (VR) collections are becoming more prevalent for use in research and as a way to discover and explore the world.¹ These technologies show great promise for a range of scholarly fields as they offer the potential for interactive visualization and analysis of artifacts, spaces, and data. Many academic libraries have developed archives for other forms of research data, but there is an absence of standards and practices for producing and managing 3D and VR content. Libraries are encountering uses for 3D and VR data and collections that require management for discovery, access, and preservation.² Following decades

Juliet L. Hardesty is a Metadata Analyst in the Library Technologies division of Herman B. Wells Library at Indiana University; email: jlhardes@iu.edu. Jennifer Johnson is Digital Scholarship Outreach Librarian in the University Library at IUPUI; email: jennajoh@iupui.edu. Jamie Wittenberg is manager of the Scholarly Communications Department of Herman B. Wells Library at Indiana University; email: jwitten@indiana.edu. Nathan Hall is Director of Digital Imaging and Preservation Services at Virginia Polytechnic Institute and State University; email: nfhall@vt.edu. Matt Cook is Digital Scholarship Program Manager in Lamont Library at Harvard University; email: matt_cook@harvard.edu. Zack Lischer-Katz is a postdoctoral fellow at the University of Oklahoma; email: zlkatz@ou.edu. Zhiwu Xie is Professor, Director of Digital Library Development in the University Libraries, and Associate Director of Smart Infrastructure Laboratory in the College of Engineering at Virginia Polytechnic Institute and State University; email: zhiwuxie@vt.edu. Robert McDonald is Dean of University Libraries, Senior Vice Provost of Online Education, and Professor of Library Administration at University of Colorado Boulder; email: rhmcdonald@colorado.edu. ©2020 Juliet L. Hardesty, Jennifer Johnson, Jamie Wittenberg, Nathan Hall, Matt Cook, Zack Lischer-Katz, Zhiwu Xie, and Robert McDonald, Attribution-NonCommercial (<https://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

of digital library experimentation, libraries have emerged as experts in digital collection creation for two-dimensional (2D) objects, developing standards for scanning, description, data interchange, web display, intellectual property, and storage/preservation. Standardization promotes data exchange and enables local entities to contribute to national portals such as the Digital Public Library of America (DPLA) and further facilitates reuse of objects and data for research in digital humanities and science. To keep up with emerging technology trends to meet the research needs of faculty, the active learning needs of students, and the curation of all types of information, it is imperative that libraries extend their expertise from 2D digital collection creation to 3D applications.

“Developing Library Strategy for 3D and Virtual Reality Collection Development and Reuse” (LIB3DVR) is a grant project³ funded by the Institute of Museum and Library Services (IMLS) to investigate these issues and provide a way forward.

The LIB3DVR project has the following four primary goals:

- Form holistic knowledge from normally siloed areas of 3D/VR by bringing together experts from a wide variety of disciplines, each with expertise in different aspects of the research lifecycle.
- Develop best practices to support 3D/VR throughout the research lifecycle, including metadata schemas; workflows and tools for capturing metadata from content creation through visualization, archiving, and reuse; and repositories that enable description, discovery, and long-term preservation.
- Establish guidelines that can serve multiple research contexts and use cases that libraries may need to support as researchers increasingly adopt 3D/VR as research tools.
- Develop strategies that libraries can use for creating policies and workflows for providing 3D/VR-related research services.

To meet the project goals, three national forums were held in different regional locales, each on a different 3D and VR theme: content creation and publishing, visualization and analysis, and repository standards and practices. Core team staff with relevant experience from partner institutions attended each forum along with expert practitioners, industry representatives, and those selected from a national call to participate.

The Washington, DC Forum on 3D/VR Content Creation and Publishing took place March 1–2, 2018, in Arlington, Virginia.⁴ The Washington, DC Forum focused on identifying necessary skills, technology, labor, and other costs of 3D/VR; workflows and metadata standards for content creation and publishing; best practices for preservation and access file formats; and implications for data curation.

The Oklahoma Forum on Visualization and Analysis took place June 13–15, 2018, in Norman, Oklahoma.⁵ The Oklahoma Forum focused on studying the integration of 3D/VR into teaching, research, and outreach. Topics included hardware and software design for developing 3D/VR tools that provide scholarly access to 3D data produced via expert modeling techniques, workflows, and abstract or numerical data sets (such as weather data).

The Chicago Forum on Repository Practices and Standards took place September 17–18, 2018, at the Big10 Conference Center in Rosemont, Illinois, to focus on 3D/VR Repository Practices and Standards. The Chicago Forum included academic librarians, scholars and researchers working in 3D/VR, digital repository developers, natural history museum practitioners, and United States government agencies involved in digital repository management and 3D/VR.

Literature Review

Challenges specific to 3D repositories have been previously explored, and various repository workflows and infrastructures have been proposed for 3D models and collections. We have identified eight projects since 2008 that have attempted to address these challenges related to 3D data. Common themes among proposals include collecting metadata centrally while keeping raw files distributed, including processing steps to generate 3D models as part of a repository workflow, and providing a way to easily access the original data for reuse and research. Flexible repository architectures and generalizable and flexible metadata models were often used, although that could be due to the experimental nature of the projects described in the literature. Some challenges were common across the literature as well and included tracking versions of 3D models and supplying reconstruction or uncertainty information about a 3D model (sometimes referred to as *paradata*).

The goal of the project described in Koller et al. is to provide a way for archaeological 3D models to be available for scholarly research and use, employing techniques to secure raw 3D data while central archival repositories collect, preserve, and track changes to 3D models.⁶ They cite digital rights management, identification of reconstruction uncertainties, version control, metadata, preservation, interoperability, and searching 3D models, along with computational analysis tools and a peer-review system for 3D models, as major challenges for cultural heritage 3D. They point to traditional scholarly journal publishing workflows as a more ideal model for how a 3D repository system should work. Similar to the way scholarly data archives work to manage research data sets accompanying published articles, the challenges involve incorporating three-dimensional capabilities for searching shapes, viewing differences in versions, supplying information about reconstruction decisions, and ensuring that digitized versions of sensitive cultural heritage sites are not compromised in their dissemination.

The 3D-COFORM project was a European-based project that ran from 2008 to 2012 and sought to “advance the state-of-the-art in 3D-digitization and make 3D-documentation an everyday practical choice for digital documentation campaigns in the cultural heritage sector.”⁷ 3D-COFORM’s Repository Infrastructure uses an API connection to client-side applications for object management and use to establish a workflow where 3D models are created and worked on over time with people located in different places using different tools to arrive at a final object.⁸ The repository is designed as a distributed content management system where raw binary files are maintained in storage at a particular organization (called a “Location”), the metadata is centralized from all Locations onto a central server, and clients, which are logged in as users, may access and edit restricted data or retrieve and download available 3D data. The system supports generating documentation (*paradata*) about processing events that produce 3D models from original scan data of different kinds (such as laser surface scanning, photogrammetry, and others) as part of the ingest process. Ingest can occur in offline mode, providing only metadata to begin, which allows large binary files to be uploaded later. It is an all-encompassing system built with cultural heritage institutions (museums, in particular) in mind to encourage effective management of objects with research-quality data included about 3D models produced and stored. Most interfaces in this project use locally installed client-side viewers.

Doerr et al. propose a repository design that integrates multiple workflows, similar to the 3D-COFORM cultural heritage project.⁹ The focus is on recording the process and tools, the provenance of activities, so that versions of the 3D models over time are available for access

and reuse. MediaHub, Alfresco, and Fedora were among the repository software systems reviewed, but the authors found "...there is no system in our knowledge, that provides an integrated high performance working environment to support the management of complex digital objects together with their significant, highly interrelated, semantic and provenance metadata information."¹⁰ The desired repository infrastructure is one that allows multiple types of users to load, access, and retrieve 3D models and all accompanying metadata using API calls to an integrated set of repositories. The object repository is viewed as a data warehouse, and the metadata repository is a semantic network representing metadata files that are also stored as objects. Querying occurs across both the object repository and the metadata repository. This design supports the following needs: the ability to search by common object shapes, document versions of objects at points in the process, and create annotations. 3D repository workflows were also conceived as centered around tracking events, using the CIDOC-CRM (conceptual reference model) standard with the CRMdig¹¹ extension.¹²

Archaeology Data Service provides different workflow recommendations tailored to the process used to create a 3D model in its *Guides to Good Practice*.¹³ For example, the recommended workflow for close-range photogrammetry produces multiple files that make up the whole of the 3D model. Each preserved deliverable should be accompanied by metadata describing the project, camera calibration, the images captured, the reference/datum for control points (if available), the model produced, and any additional processing that was done to create these files.¹⁴

For an experimental project to store archaeological 3D models of Uchi Maius, a site of Roman and Islamic remains, Felicetti and Lorenzini reviewed the 3D-COFORM Distributed Object Repository and the Fedora repository in 2011 and decided on using Fedora as their 3D repository. The repository was combined with a Solr index and open source tools for ingest and annotation, custom constructed for this use case. While this system worked for their proof-of-concept project, they cited the need for more "...detailed and high-quality metadata" and improved preservation functionality to support the needs of 3D collections.¹⁵

MorphoSource, an online 3D data archive from Duke University, is attempting to make "high-fidelity scan data as widely accessible as possible."¹⁶ MorphoSource is based on open source software (Linux/Apache/MySQL/PHP or LAMP stack) using CollectiveAccess,¹⁷ an open source media collection management application as its code base. This software stack, while open source, is unique, and work is underway to remake MorphoSource using the Samvera (formerly Hydra) software stack¹⁸ along with Duraspace's Fedora repository¹⁹ to align more with a wider open-source community implementation and library archival standards and improve the preservation side of the application. MorphoSource allows upload and download of a variety of 3D models or data sets with a registered user account to track activity. These 3D models are based on physical real-world biological specimens or parts of biological specimens, with the aim of only representing these physical objects and not, for instance, composite specimens created from multiple physical specimens. MorphoSource stores both raw data and derivative data; however, the documentation on the site does not define these terms. Downloads have access controls available and are always in compressed ZIP file format. For end users, these files can be too large to easily unpack and the contents are unknown until that unpacking occurs. Reducing the size of raw data before upload is recommended as a best practice by MorphoSource to help ease upload and download times and the use of the data, so an ingest process that can assist in those activities would be beneficial to a 3D repository

system. MorphoSource also reports that long-term preservation is still an unresolved concern because 3D images and metadata do not yet have accepted preservation standards.

The CARARE project (Connecting ARchitecture and ARchaeology in Europeana) has made recommendations for aggregating and publishing 3D objects in PDF format with accompanying metadata that conforms to the Europeana Data Model.²⁰ The 3D PDF digital resources remain with the content provider in their repository while the metadata for those resources are aggregated using the CARARE metadata schema, which is then mapped to the Europeana Data Model.²¹ The CARARE repository serves only to store and process metadata, and it uses the Fedora digital repository as its underlying architecture. CARARE metadata include information about the physical monuments, activities conducted, and digital resources produced. 3D is included in the digital resources but CARARE is not limited to 3D metadata. Searching Europeana now, there does not appear to be any 3D content available, although a task force has been recently formed to address this area.²²

In exploring the concept of data sustainability and accessibility, the MayaArch3D project, which built a 3D virtual environment for analyzing archaeological data,²³ encountered four key issues with regard to 3D data: data security/sensitivity; web-based dissemination; conveying uncertainty; and data storage, reuse, and peer review.²⁴ In their project, security is addressed by providing different levels of user access with sensitive location-based data protected with additional restrictions. To address web-based dissemination challenges (such as high data transfer rates required for high resolution data types), the project made use of JavaScript libraries and JSON file formats to provide lightweight online access to massive amounts of data. The researchers also incorporated techniques to show a limited view (such as by loading landscapes one tile at a time), while keeping the entire data set (that is to say, the terrain of the landscape) available. Conveying uncertainty about construction of 3D models, particularly in architecture and archaeology, has often been accomplished using visual indicators of some kind to express modeling certainty,²⁵ but the MayaArch3D project accompanied visual indicators with links to textual explanations (paradata) and metadata explaining how, when, and by whom the models were created. The authors contend that this provides more information for all viewers but can especially help in the peer-review process of a published 3D model. To address data storage and reuse, the data sets were placed in a data repository in Germany that specializes in archaeological data, including 3D. DOIs were generated for these objects, and the metadata was mapped to CIDOC-CRM for better interoperability. The viewers and tools developed were also made open source and reusable by other researchers with their data sets to encourage further use and development.

This literature review reveals a number of unaddressed issues facing the field of 3D/VR:

- There is no existing repository solution that effectively manages 3D models and collections in standardized ways (approaches are *ad hoc* and customized to the particular project). A repository that can manage 3D models will be most effective if it can understand the different types of file outputs and metadata needs for different 3D processes.
- Online access to discover and reuse original 3D data for research purposes is a challenge due to the size of 3D models and VR environments.
- The lack of standards or best practices leaves long-term preservation needs unfilled and preservation activities undefined or sketchy at best.

More discussion within the community is necessary to establish best practices and standard workflows around 3D repositories. The third LIB3DVR forum meeting in Chicago convened to review and discuss how this might be achieved.

This paper addresses the following research questions that were the focus of the Chicago Forum:

Q1: What repository features are necessary or helpful for 3D data management?

Q2: What workflows or best practices can be applied to 3D data to aid its storage and management in information institutions?

Q3: What are the implications of 3D data on current preservation models?

Methods

The forum meeting occurred during a period of a day and a half and was attended by 16 invited participants across academic disciplines (art, anthropology, sciences), academic libraries, museums, the repository software industry, and a scientific government agency. Participants were selected by identifying national experts in representative fields, with an eye toward achieving institutional, disciplinary, racial, and gender diversity. The organizers shared an initial participant list with an advisory board that provided further input on the selection of participants, and an email was sent to the selected experts, inviting them to participate in the forum. The assembled participants provided a breadth of experience and knowledge in 3D/VR and repository practice.

The forum meeting began by reviewing national repository/preservation efforts, which informed part of the environmental scan in this article. The forum then employed a nominal group technique to explore key challenges in 3D/VR repository management.²⁶ Nominal group technique is a consensus-building method for achieving general agreement on a topic through face-to-face small group discussions. It is useful for understanding stakeholder views. The conveners arranged the participants in small groups for discussion, organized as much as possible to keep a mix of various represented areas within each small group, and constructed activities based on agile project design concepts used for rapid prototyping.²⁷ The smaller groups then reconvened in the larger forum to report and confer on their discussion findings, developing ideas for better workflows and practices. In the first small group discussion, common terminology was offered by facilitators to help frame the overall forum meeting discussion, and then an activity was conducted to identify and prioritize high-level features and functions necessary to support 3D/VR models in repository systems. The identification and prioritization activities were conducted using sticky notes on whiteboards as a way to record and easily order identified features and functions.

Community Standards for 3D Data Preservation (CS3DP) is a separate IMLS-funded project “bring[ing] together stakeholders to work toward establishing shared standards, practices, and technologies for digital 3D data preservation, documentation, and dissemination.”²⁸ The Chicago Forum used the CS3DP Intergroup Terminology Glossary as reference for 3D/VR terminology to establish a common language to support clear communication between participants from different backgrounds. Additionally, the repository standards and practices conversation was aided by agreement on the following definitions:

- Practices: ways of doing things in an institution
- Policies: documented practices
- Standards: shared and ratified policies

- Repository: digital object management platform that does any combination of the following: publishes, provides access to, and/or preserves output
- Storage: digital space where data resides (local or cloud-based servers); may be local or remote, with regard to a repository

The ensuing discussions within the forum confirmed the suitability of these definitions, and participants did not request any subsequent revisions to them. Repository standards and practices were discussed as they apply across multiple institutions, but the ways in which repository systems are used for storage, access, and research focused on individual institution implementations as use cases, particularly as features and workflows were considered. The CS3DP Intergroup Terminology Glossary was a useful tool in communicating about repository standards and implementations across organizations.

The full group of participants then reconvened and considered case studies and workflow diagrams from existing 3D/VR repositories to see examples of systems designed specifically to store and manage 3D/VR models and to consider how workflow diagrams can be constructed. The small groups were tasked with creating a new workflow diagram, drawn on paper and using sticky notes if needed, that considered the high-level features and functions identified in the previous activity and incorporating any additional features identified in the case studies. The full participant group discussed the workflow diagrams to compare and evaluate the features and workflow concerns discussed in the small groups.

Once the repository features were prioritized and the workflow diagrams reviewed, the full group and small groups discussed the implications of what was produced through these activities. The topics discussed by the full group included the following implications:

- Implications for different levels of library support
- Implications for existing digital preservation models and practices

The small groups held one final discussion on implications for 3D/VR creation and publication (based on findings from the LIB3DVR forum in Washington, DC) and 3D/VR visualization and analysis (based on findings from the LIB3DVR forum in Oklahoma). Finally, the full group convened to conclude by identifying areas in need of further work and additional stakeholders who should be involved in future discussions about repositories for 3D/VR collections.

The researchers synthesized the findings of the forum and disseminated an early draft to participants to validate the findings. Virginia Tech's Institutional Review Board approved the research protocol (IRB #18-770).

Environmental Scan

To help frame discussion of 3D/VR repository standards and practices at the Chicago Forum meeting, the forum group reviewed national preservation efforts and repository services in terms of their capabilities to handle 3D models and VR environments. The NIH 3D Print Exchange is a strong example of a platform that is gathering 3D models from various sources and providing viewing and download access. The service is based on the free and open-source Drupal content management system and includes a custom module to convert 3D data to a printable model. Viewable models are provided in X3D, an ISO-ratified open standard file format for 3D scenes and objects.²⁹ Currently, the NIH 3D Print Exchange holds approximately 1.5 terabytes of data. The main uses for this service include 3D printing medical equipment (such as prosthetics and assistive devices) and bacteria and virus specimens.

The group also reviewed the Native American Graves Protection and Repatriation Act (NAGPRA), which allows tribes to regain artifacts that have been placed in museums and other cultural heritage institutions, and the use case of a biologist studying human remains via computed tomography (CT) scans. Ethical questions surround whether or not items being repatriated can be digitized as 3D models before returning them and how best to approach digitizing human remains for research use. It was found that there are no standards, laws, or guidance for creating and managing these kinds of 3D digital collections.

APTrust,³⁰ MetaArchive,³¹ and the Digital Preservation Network (DPN) were also reviewed as examples of national preservation efforts.³² While these efforts vary between commercial and noncommercial, they do provide common methods for data storage (via tape or cloud services) and digital preservation practices to provide trustworthy backup repositories for long-term preservation. A common theme that is still problematic for preserving 3D models and VR environments is storage limitations and cost. As an example, before it announced its sunsetting, DPN required \$2,000 per terabyte, and Indiana University-Purdue University Indianapolis, an institution just beginning to gather a 3D digital collection, is already at almost 2 terabytes of data with only a few hundred objects digitized. The scale for preserving 3D models and VR environments is much larger than the national efforts that are providing standards for depositing and replicating digital collections of other formats of data.

Findings

Q1: Repository Features for aiding 3D/VR Data Management

Using national repository and preservation efforts as examples, small groups identified high-level features and functions necessary to support 3D/VR models in a repository system. Groups noted that while current repositories provide the necessary functions, such as assigning metadata and fixity checking, considerations need to be reviewed for incorporating 3D/VR files into these existing systems. They identify six areas that need to be addressed: searchability; security, ownership, and licensing; streaming capability; download capability; scalability; and preservation.

Searchability

While digital repository systems provide search capabilities, the small groups included faceted searching as a necessary high-level feature. It was noted that, even at a basic level, keywords and tagging need to be a feature that is included for searchability. These types of search features are important for 3D data because, similar to other non-text-based digital media, content identification in the form of keywords, tags, subjects, and genres is needed separately from the 3D models themselves for discovery and research use in order to be effective. Additionally, VR environments can benefit from search functionality within the viewing interface and offer a different experience from the more traditional "Search Within" capabilities included in text-based digital content. Within VR environments, searching can be immersive and focus on identifying objects and metadata associated with those objects as they appear within the virtual context of the environment. Selecting items to find out more about them or searching and finding results within the visual field of the virtual environment resembles an activity that is more similar to videogame functionality than to traditional "Search Within" functionality in search interfaces that produce search results with highlighted text portions.

Security, Ownership, and Licensing

Within many organizations, security measures and intellectual property rights need to be considered within the digital repository. These measures include determining ownership and licensing; controlled access; and determining how content can and will be distributed by accessing a data use agreement. Such agreements can vary across content, disciplines, and use types. For example, when levels of access were discussed by the forum participants, emergent themes included the importance of protecting the cultural heritage of indigenous communities and the portrayal or representation of those communities through 3D content and experiences. Participants identified potential solutions for managing such content, citing Mukurtu (a content management system built with indigenous communities specifically for digital cultural heritage management) as an example of how to manage access in conformance with indigenous cultural protocols.³³ Issues of ownership and licensing potentially delve beyond a repository management system if the digitized items involve human beings or are part of collections taken from historically oppressed communities.³⁴ The degree to which the scans themselves are copyrightable is somewhat open to debate, although the general concept of scanning only for reproduction purposes likely does not produce a copyrightable digital object.³⁵

Streaming Capability

Providing access to an interactive 3D viewer in a web browser or embedded in the repository platform display interface was seen as a functional requirement in the participant discussions. This feature allows interaction and exploration of objects, providing ease of use and access, without having to download a full 3D file for use in client software. Maintaining the performance of a web-based viewing platform was discussed as one area of concern, and simplified models may need to be generated specifically for online and/or streaming distribution of 3D content. The International Image Interoperability Framework (IIIF) offers a possible avenue for exploring standardized streaming access through the development of specifications to support 3D use cases.³⁶

Download Capability

3D/VR is created for various disciplines, users, and applications. Download availability will vary based on the institution and there is a lack of interoperability standards limiting streaming, viewing, and archival-quality downloads. Nevertheless, the small groups identified the ability to download files as a key feature within a digital repository, one that needs to be able to provide a higher resolution version for archiving and for research activities. Associated metadata was also discussed and may take the form of a separate file. For example, the metadata used for retrieval may be required for indexing in a repository, but a different set of metadata elements for using 3D models for research purposes may be stored in an accompanying file (sometimes referred to as a “sidecar” file). Finally, participants discussed the importance of considering file format choices and standards, since the usefulness of certain content (or complete experiences, in the case of VR applications) may be undermined by discipline-specific or proprietary file types.

Scalability

Storing and serving 3D collections introduces issues of scale that strain or might go beyond currently available repository technologies. Supporting workflows to manage an entire collection of 3D objects or VR environments composed of very large files is essential to having a

usable repository system. Asynchronous storage, in which repositories allow management of digital objects (metadata and structure), while the actual media is stored and called up from separate servers that can handle the large storage size or streaming bandwidth necessary to serve various output versions (for access, research, and preservation purposes) is an example of the type of repository system that might be necessary for managing 3D/VR collections. Cloud-based services offer possible capacity if they can be implemented for use with a repository management system that offers the other features identified here.

Preservation

Standard Format

Multiple groups included standardized file formats as a necessary high-level repository feature. Where repository systems and end user delivery interfaces might have predefined file format expectations for digital objects such as images, text, video, or audio, those same systems are not opinionated about 3D/VR file formats nor do they have the ability to understand those formats for access and viewing purposes. Community-agreed standards would advance the ability of repositories to preserve 3D/VR objects, making them usable, manageable, and shareable into the future.

Versioning

There are many formats and file types that need to be considered within the 3D/VR environment, and 3D workflows are complex processes that can produce many different file versions. It is important to document the creation and management of each release of a 3D file, for instance, and version for a VR environment. Considerations might include identifying whether a 3D/VR object has been optimized for printing versus visualization in a VR environment. Versions for different uses and versions of the same file are both considerations for 3D/VR models. Having capacity to track version changes over time and create multiple versions of the same object for different uses are features supported by repository management systems for other media formats, but the scale required for versioning 3D/VR models might prove challenging.

Data Scrubbing

3D techniques and workflows such as those associated with computed tomography (CT) used in medical imaging and other fields can involve personally identifying information in scans involving human beings. This sensitive data must be scrubbed to protect personal (such as patient) privacy. Repository management systems and workflows would benefit from tools to identify and scrub personally identifiable data from medical imaging or other types of sensitive data involving humans.

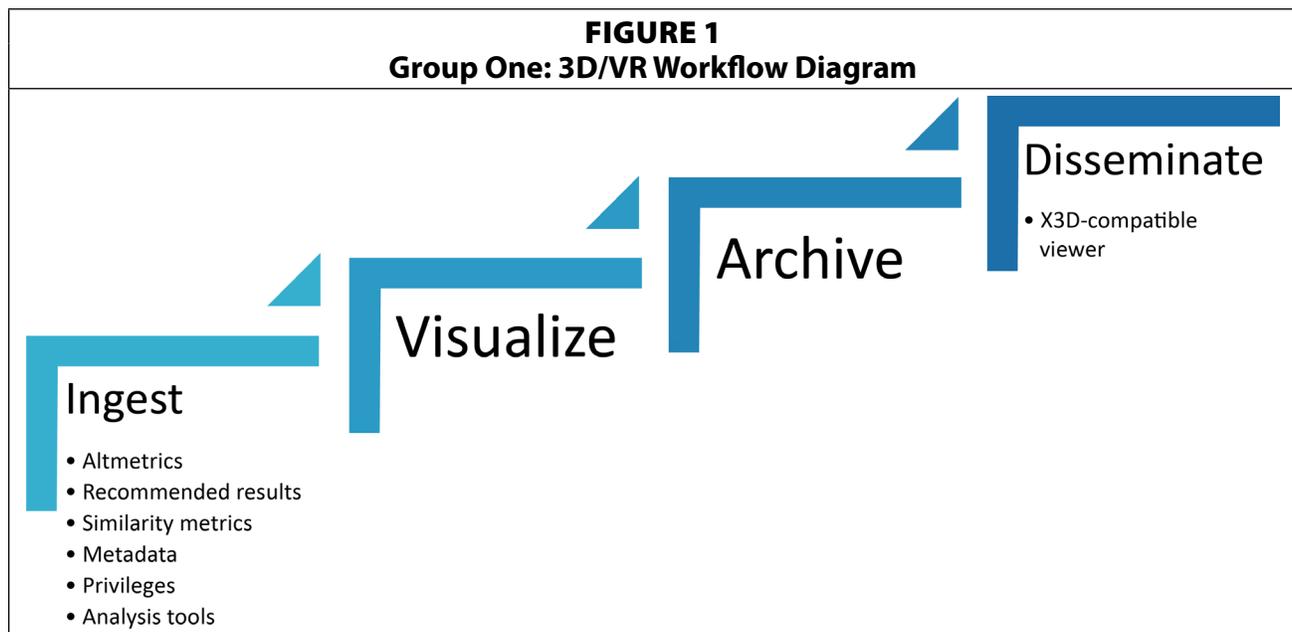
Dependencies

In the case of VR experiences, myriad dependencies—including associated hardware, software, and network components—need to be considered if long-term preservation (and potential reproduction) is the goal. For example, within a given virtual environment, certain files or lines of code determine how light will appear to reflect off objects in a consistent manner, and there might be different pieces of code controlling this aspect depending on whether the user is exploring an exterior space or an interior space in the virtual environment, since exterior spaces

should mimic natural light and interior spaces should mimic artificial light coming from sources such as lighting fixtures or windows. This facilitates a more natural and realistic experience within the environment. If the code that controls these aspects is not preserved, then the virtual environment experience will not appear as originally intended. The added level of complexity associated with VR applications, as opposed to discrete 3D objects, led participants to conceptually differentiate 3D and VR outputs in terms of their long-term preservation needs. While 3D and VR are interrelated, VR was seen as more complex and thus more difficult to preserve.

Q2: Workflows and Practices for 3D Data Management

Once high-level features and functions were identified, the three small discussion groups were tasked with creating workflow diagrams, producing three examples, which appear as figures 1–3 below.



Group One identified four components needed to incorporate 3D/VR objects into a repository (see figure 1, above). These components include **ingestion, visualization, archiving, and dissemination**. The ingestion component included the concept of middleware microservices to help in processing files associated with 3D models and VR environments before ingestion occurs. For dissemination, the group discussed X3D³⁷ as a standardized file format for researcher access and expressed the importance of a standardized file format-compatible viewer.³⁸ Existing 3D viewers were also reviewed and found to be consistently lacking necessary functions, including: usage metrics, an analysis package, and tools specific to measurement. These features are necessary within many institutions specific to research needs.

Group Two created a workflow diagram (see figure 2, below) that outlines a framework based on a scholarly data publishing use case (that is, depositing a publishable 3D model), with the goal of creating and/or extending the work of an already existing digital library workflow. The workflow is based on the OAIS reference model.³⁹ The group noted the need for a user-interface (UI) layer. In addition to the UI layer, it was noted that an API layer could be included for computational access. However, inclusion of an API may not always be possible or suitable

FIGURE 2
Group Two: 3D/VR Workflow Diagram

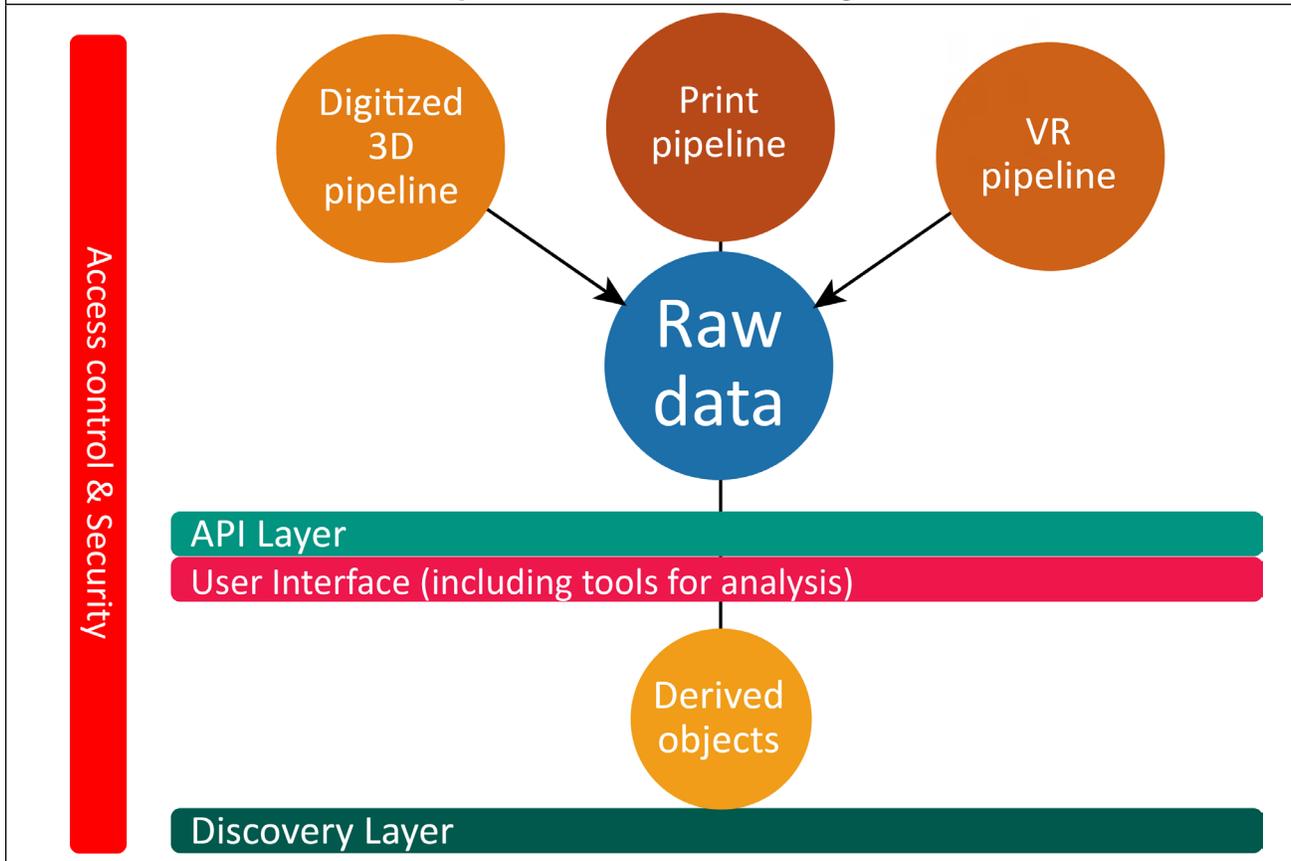
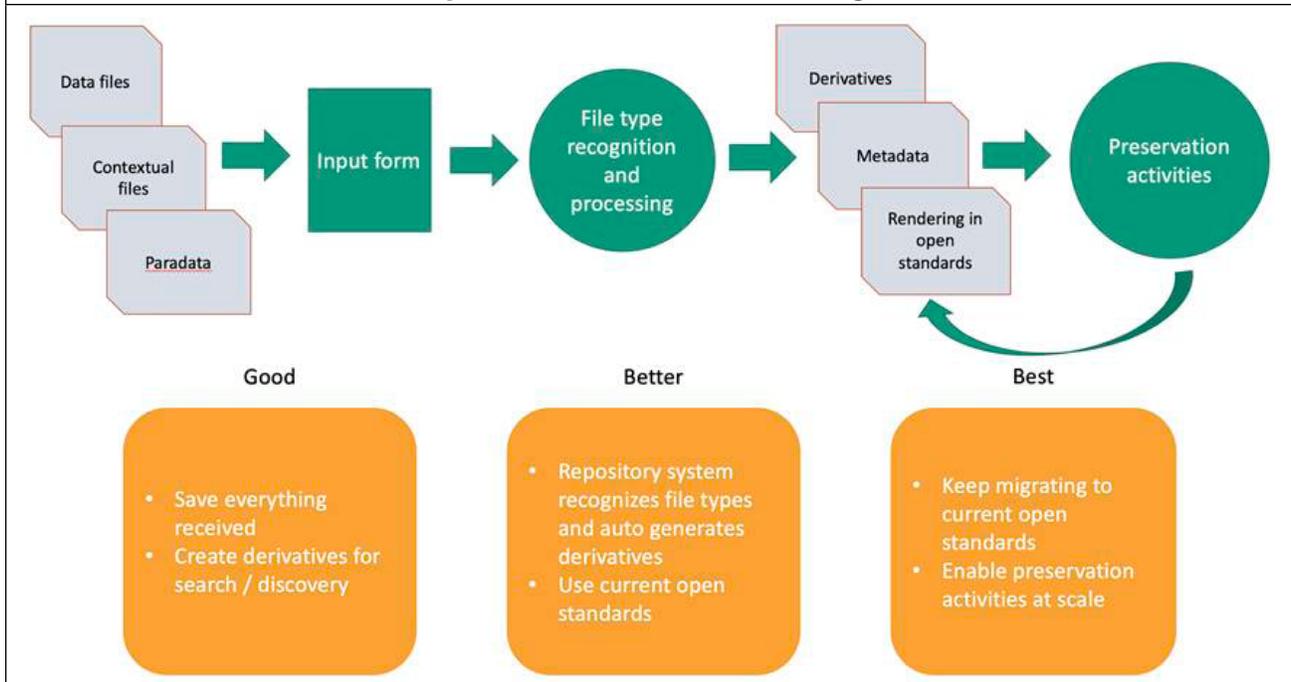


FIGURE 3
Group Three: 3D/VR Workflow Diagram



depending on the business needs of the organization. Technical resources, policy restrictions, and data ownership concerns may all be factors in determining whether an API should be included in a workflow for 3D data in libraries. This workflow would also take into consideration various uses of the files for educational and research purposes. Multiple derivatives of the original files would be generated on demand as needed, and persistent identifiers in the form of minted DOIs could be generated to track derivatives and downloads as remixed or adapted objects. When it exists, linked open data would be added to relate the 3D/VR data to the original, physical object.

Group Three acknowledged that existing repository platforms already provide many features and functions that are necessary to support 3D/VR files but that they need to be scaled up. They outlined the basic workflow process and suggested three levels of scalable preservation—“good,” “better,” and “best”—with each level offering more automation and long-term preservation services to support bigger files and collections (see figure 3, above).⁴⁰ The group identified the following considerations for further exploration:

- Examine the basic parts of the repository system and identify what is unique about 3D objects and how users need to interact with them, including processing input files before ingestion and generating derivatives for access on demand.
- Identify what data files and contextual files need to enter the repository system.
- Identify what type of file recognition system is needed.
- Consider preservation activities and goals to determine which files are important for long-term preservation. Ideally this would incorporate standardized file formats, but more realistically it might be the most complete version or the best version for file use. For example, computer-aided design (CAD) is a format that generally requires rendering. In this case, the rendered stereolithography (STL) file is an important part of the archival package.

Q3: 3D Data in the Context of Preservation Models and Practices

The high-level features and possible workflows identified by the forum meeting participants considered through the lens of different preservation models and practices revealed implications for design and implementation. In particular, the Digital Curation Centre’s (DCC) Digital Curation Lifecycle,⁴¹ Open Archival Information System’s (OAIS) reference model (ISO 14721:2012),⁴² and Force11’s Findable Accessible Interoperable and Re-usable (FAIR) Data Principles⁴³ were discussed at the forum meeting in connection to the outputs created by the participants. Mapping the forum-generated workflows to widely adopted repository models was an essential exercise in appraising the feasibility of supporting 3D/VR in existing library repositories. Participants identified an additional repository model that could benefit libraries seeking to expand support for 3D/VR: E. Zierau’s 2017 paper describing the Outer OAIS–Inner OAIS Model for Distributed Digital Preservation. This model “is an overlay enabling the description of distributed digital preservation systems in a way that conforms to the OAIS Reference Model.”⁴⁴

The DCC Lifecycle model (see figure 4, below) represents a set of actions and tasks that are necessary to effectively curate and preserve digital content from conceptualization to disposition. Full lifecycle actions (Description and Representation Information, Preservation Planning, Community Watch and Participation, Curate and Preserve) along with Occasional Actions (Dispose, Reappraise, Migrate) were accepted by the group as useful models for the preservation of 3D/VR. Regarding the sequential actions detailed by the model (Concep-

FIGURE 4
DCC's Digital Curation Lifecycle⁴⁵

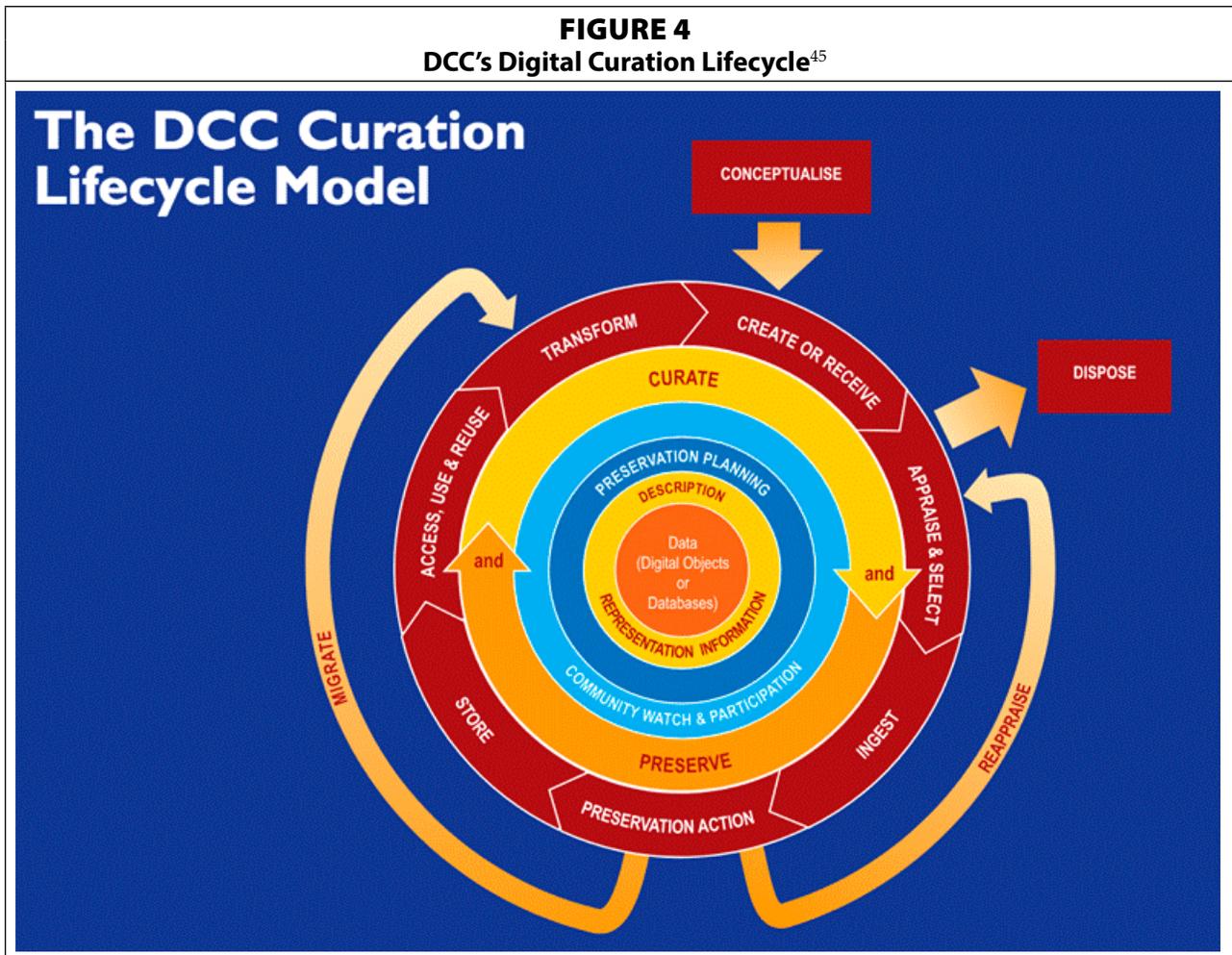
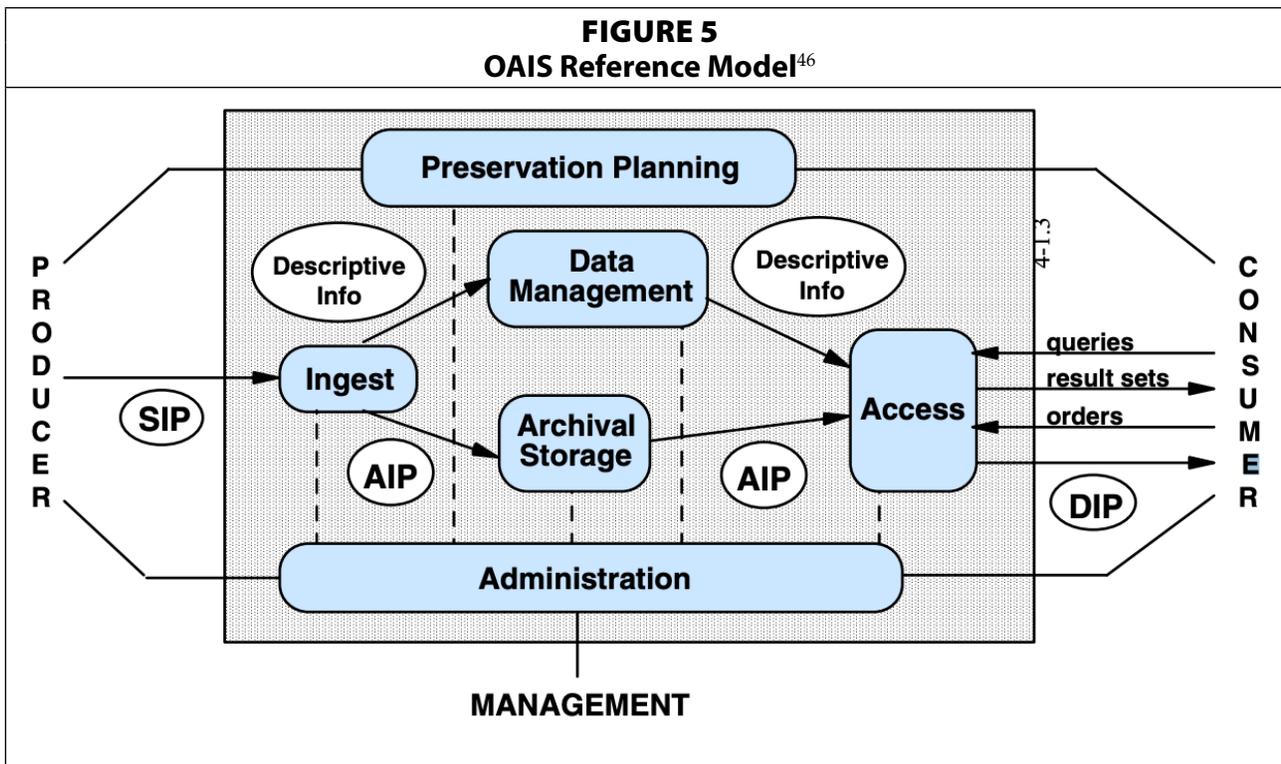


FIGURE 5
OAIS Reference Model⁴⁶



tualise; Create or Receive; Appraise and Select; Ingest; Preservation Action; Store; Access, Use and Reuse; Transform), participants expressed that the unique nature of 3D/VR content may require a modification of the DCC Lifecycle model. Participants noted the proprietary nature of generating 3D models, often created using commercial software, and emphasized that transforming data (such as using photogrammetry software) repeatedly does not always produce the same results. In the case of VR, participants shared that it is difficult to decouple the digital content from the hardware requirements and that the necessity of capturing hardware and environment is not adequately represented in the model. Participants concluded that applying 3D data lifecycles to this model might create some blockages in the sequential paths shown, due to lack of transparency and openness associated with file processes.

The OAIS Reference Model (see figure 5, above) is a conceptual framework that enables organizations to design, describe, and compare the design and deployment of digital information management systems. The model defines three types of information packages that move through the system: Submission Information Package (SIP), Archival Information Package (AIP), and Dissemination Information Package (DIP). In the case of 3D/VR, these packages can be mapped to systems as well as to users or stakeholders, where the creator of the object submits the SIP, the data manager assembles the AIP, and the reader/patron accesses the DIP. Assessing the usefulness of this model for 3D/VR gave rise to concern about limitations in the framework for dissemination. Particularly, there is a difference between experiencing the 3D/VR medium as an immersive and interactive environment and downloading a package for reuse. These are both types of access but the requirements for accessing/viewing and for downloading/reusing may necessitate very different types of digital objects from different storage sources. The overall process the OAIS model suggests is applicable to 3D/VR, but it requires a more nuanced view of access for the different ways in which 3D/VR might need to be delivered.

Force11's FAIR Data Principles

The FAIR Data Principles, published in 2016, are a series of guiding principles that aim to improve the findability, accessibility, interoperability, and reproducibility of data. There have been attempts to map these principles directly to repositories.⁴⁷ Dunning et al. found in their analysis of more than 40 data repositories that implementation of FAIR principles can be challenging, especially in terms of the interoperability and reusability principles.⁴⁸ However, they concluded that adhering to the FAIR principles often improved repository compliance with the various FAIR facets within each principle. These principles were the most controversial of the models presented at the forum. Participants reiterated that interoperability is a challenge due to the many proprietary technologies involved in developing models. Participants questioned whether Interoperable Metric I1 “(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation” is achievable. The forum concluded that it may be achievable for metadata about a 3D model but likely is not for the 3D data itself. There were questions about agreement on different kinds of 3D data (for instance, the term “mesh”) and what files are and are not necessary to constitute a 3D model. Individuals and communities disagree on formal definitions of terms for 3D/VR vocabulary. For example, forum participants who work with medical CT scanning have different vocabularies from participants working with industrial CT scanning. Participants suggested that this could be an area that benefits from “better” rather than “best” practices. Additionally, the forum discussed the extension of

the FAIR principles to “FAIR-EST” in order to represent the importance of Ethical, Sustainable, and Transparent practices in 3D and VR. The forum also discussed distinguishing between a playable or instantiable model versus all data associated or comprising a model and that a playable object might need to be a discrete metric when considering the FAIR principles.

Discussion

The findings from this forum suggest that collection management for 3D/VR models is likely to be tightly coupled to creation workflows and possibly even use workflows. To effectively archive a 3D/VR model, it might require that specific pipelines be created for managing particular types of 3D/VR models created by particular software within a single repository system. While original authoring software is not necessary for accessing and viewing 3D/VR models, without standards for file types the software used to create 3D/VR models would need to be either part of a repository system, or the software, and the operating system that it runs on, would need to be included in the archival package to ensure preservation. It might also be necessary to include viewing software as part of the repository system or archival package depending on the 3D/VR model being preserved (if, for example, a proprietary viewer is necessary for basic access and viewing of the model). While this limits the capacity for accepting various 3D/VR outputs, a recommendation like this might also encourage libraries to enact common standards regarding which 3D/VR file formats they will accept into their repositories and develop trustworthy procedures for normalizing or converting noncompliant file formats. It has the potential to encourage standardization beyond libraries as well.

3D data also impacts long-term preservation practices and how libraries should consider managing 3D/VR collections in the long term. National digital preservation efforts to date show limitations in scale and cost. For example, DPN announced in 2018 that they would cease operations. Library capacity and audience might impact the best support path. If a library needs to support a research use case involving publishing and reproducibility, then capacity and workflow is needed for downloading the complete model and ensuring long-term preservation of that model by actively managing and migrating formats or bringing along the software environments needed to reuse a proprietary model (software/hardware emulation or migration). Repository systems that manage and store access-only versions of a 3D object or VR environment might prove most feasible for libraries only supporting access for discovery and viewing and might aid aggregation efforts for preservation such as DPN. If an aggregation effort still proves too costly for the derivative form of an object due to file sizes, a model similar to the Digital Public Library of America’s aggregation model⁴⁹ that gathers metadata and points back to the original digital object at its home institution offers opportunities to increase access points while limiting the impact of the space needed for these digital objects across the community.

Conclusion

By reviewing current national repository and preservation efforts, LIB3DVR Chicago Forum participants recognized the importance of analyzing current repository features; examining workflows and best practices; and outlining implications of 3D data on current preservation models. The LIB3DVR project is producing a roadmap for libraries that will provide recommendations for incorporating 3D/VR collections into the workflows of creation, use, and management that libraries provide for other types of digital collections. This roadmap will

consider different levels of support that libraries might commit to for supporting 3D/VR collections and will be based on audience needs, such as a focus on research use or discovery and access only.

This meeting helped to clarify some of the issues specific to storing and managing 3D models and VR environments. High-level features identified issues of scale, capacity, and access unique to 3D/VR collections. The constructed workflow diagrams produced by forum participants offer visions of implementation that emphasize the need for standard file formats and common workflows to enable an entire set of 3D/VR models to be manageable together. Finally, reviewing these features and workflows against current preservation models revealed the disparity between current best practice for preservation, where packaging involves a high level of standardization, and the current landscape of 3D/VR creation that involves a variety of proprietary systems, formats, and outputs with limited usefulness for research purposes, especially in terms of their capabilities for transparency and data reuse.

While the forum findings suggest ways to move initiatives forward, work is still required to address the following areas:

- Intellectual property and fair use expertise are still needed to inform managing 3D/VR use.
- Public libraries are often not managing repository systems on their own. Providing access to 3D/VR collections from aggregators or other providers is in their purview and merits further feedback from their viewpoint of how best to manage that access.
- News media and other industries also creating and working with 3D/VR collections might provide insights applicable to library considerations. These additional inputs along with reviewing the Outer OAIS–Inner OAIS Model for Distributed Digital Preservation would offer a more complete picture of the best practices and standards for managing 3D/VR collections.

Acknowledgments

This project was made possible in part by the Institute of Museum and Library Services (LG-73-17-0141-17). The views, findings, conclusions, and recommendations expressed in this program do not necessarily represent those of the Institute of Museum and Library Services. We would also like to thank our participants (John Bates, JP Brown, Margaret Dolinsky, Heather J.H. Edgar, Darrell Hurt, Jamie Kelly, Meghan McCarthy, Jaime Mears, Jennifer Moore, April Isch Neander, Mark Phillips, Helen Robbins, Stefan Serbicki, Kate Webbink, and David Wilcox) for their time, contributions, and cooperation in this study. Finally, we would like to thank our advisory board: Bryan Carter, University of Arizona; Margaret Dolinsky, Indiana University; Carl Grant, University of Oklahoma; Patricia Hswe, Mellon Foundation; Jason Jerald, NextGen Interactions; Fred Limp, University of Arkansas; Clifford Lynch, CNI; Jerome McDonough, University of Illinois; Jennifer Moore, Washington University at St. Louis; and Diane Zorich, Smithsonian.

Notes

1. Susan Jang et al., “Direct Manipulation Is Better Than Passive Viewing for Learning Anatomy in a Three-dimensional Virtual Reality Environment,” *Computers & Education* 106 (2017): 150–65, <https://doi.org/10.1016/j.compedu.2016.12.009>; Andries van Dam, David H. Laidlaw, and Rosemary Michelle Simpson, “Experiments in Immersive Virtual Reality for Scientific Visualization,” *Computers & Graphics* 26, no. 4 (2002): 535–55, [https://doi.org/10.1016/S0097-8493\(02\)00113-9](https://doi.org/10.1016/S0097-8493(02)00113-9).

2. David Koller, Bernard Frischer, and Greg Humphreys, "Research Challenges for Digital Archives of 3D Cultural Heritage Models," *ACM Journal on Computing and Cultural Heritage* 2, no. 3 (2009): 7:1–7:17, <http://doi.acm.org/10.1145/1658346.1658347>; Manjula Patel et al, "Metadata Requirements for Digital Museum Environments," *International Journal on Digital Libraries* 5, no. 3 (2005): 179–92, <https://doi.org/10.1007/s00799-004-0104-x>; Zack Lischer-Katz, Matt Cook, and Kristal Boulden. 2018, "Evaluating the Impact of a Virtual Reality Workstation in an Academic Library: Methodology and Preliminary Findings," presented at the ASIS&T Annual Meeting, Vancouver, BC, Canada, <https://shareok.org/handle/11244/317112>.
3. See <https://www.ims.gov/grants/awarded/lg-73-17-0141-17> for a description of the grant awarded.
4. For a report on findings from this forum, see Nathan Frank Hall et al., "Challenges and Directions in 3D and VR Data Curation: Findings from a Nominal Group Study," *International Journal of Digital Curation* 14, no. 1 (2019): 150-167, <https://doi.org/10.2218/ijdc.v14i1.588>.
5. For a report on findings from this forum, see Matt Cook et al., "Challenges and Strategies for Educational Virtual Reality: Results of an Expert-Led Forum on 3D/VR Technologies Across Academic Institutions," *Information Technology and Libraries* 38, no. 4 (2019): 25-48, <https://doi.org/10.6017/ital.v38i4.11075>.
6. Koller, Frischer, and Humphreys, "Research Challenges for Digital Archives of 3D Cultural Heritage Models."
7. Ralf Sarlette et al., "3D-COFORM: Tools and Expertise for 3D Collection Formation," available online at <http://cg.cs.uni-bonn.de/en/projects/3d-coform-tools-and-expertise-for-3d-collection-formation/> [accessed August 19, 2019].
8. Xueming Pan et al., "A Repository Infrastructure for Working with 3D Assets in Cultural Heritage," *International Journal of Heritage in the Digital Era* 2, no. 1 (2013): 143–66.
9. Martin Doerr et al., *A Repository for 3D Model Production and Interpretation in Culture and Beyond* (VAST10: The 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, 2010), <http://dx.doi.org/10.2312/VAST/VAST10/097-104>.
10. Doerr et al., *A Repository for 3D Model Production and Interpretation in Culture and Beyond*, 98.
11. For more information on this extension, see <http://www.cidoc-crm.org/crmdig/home-2>.
12. Martin Doerr et al., *A Framework for Maintaining Provenance Information of Cultural Heritage 3D-Models* (London, UK: Electronic Visualisation and the Arts EVA, 2014), 268, <http://dx.doi.org/10.14236/ewic/EVA2014.63>.
13. Adam Barnes and Kieron Niven, "Guides to Good Practice," *Archaeology Data Service / Digital Antiquity*, available online at <http://guides.archaeologydataservice.ac.uk/g2gp/Main> [accessed August 19, 2019].
14. Adam Barnes and Kieron Niven. n.d. "Close-Range Photogrammetry: A Guide to Good Practice." *Archaeology Data Service / Digital Antiquity*, available online at https://guides.archaeologydataservice.ac.uk/g2gp/Photogram_Toc [accessed May 29, 2020].
15. Achille Felicetti and Matteo Lorenzini, "Metadata and Tools for Integration and Preservation of Cultural Heritage 3D Information," *Geoinformatics FCE CTU Journal* 6 (2011): 124, <https://doi.org/10.14311/gi.6.16>.
16. Doug M. Boyer et al., "Morphosource: Archiving and Sharing 3-D Digital Specimen Data," *The Paleontological Society Papers* 22 (2017): 158, <https://doi.org/10.1017/scs.2017.13>.
17. See <https://collectiveaccess.org/>.
18. See <https://samvera.org>.
19. See <https://duraspace.org/fedora/>.
20. Emilia Masci et al., "3D in the CARARE Project Providing Europeana with 3D Content for the Archaeological and Architectural Heritage: The Pompeii Case Study," *Proceedings of 18th International Conference on Virtual Systems and Multimedia* (2012): 227–34.
21. Masci et al., "3D in the CARARE Project Providing Europeana with 3D Content for the Archaeological and Architectural Heritage," 228.
22. See <https://pro.europeana.eu/project/3d-content-in-europeana>.
23. See <https://mayaarch3d.org/en/>.
24. Heather Richards-Rissetto and Jennifer von Schwerin, "A Catch 22 of 3D Data Sustainability: Lessons in 3D Archaeological Data Management & Accessibility," *Digital Applications in Archaeology and Cultural Heritage* 6 (2017): 38–48, <https://doi.org/10.1016/j.daach.2017.04.005>.
25. See projects and methods cited in Richards-Rissetto and von Schwerin, "A Catch 22 of 3D Data Sustainability: Lessons in 3D Archaeological Data Management & Accessibility," 44.
26. Sara S. McMillan, Michelle King, and Mary P. Tully, "How to Use the Nominal Group and Delphi Techniques," *International Journal of Clinical Pharmacy* 38, no. 3 (2016), 655–62, <http://dx.doi.org/10.1007/s11096-016-0257-x>.
27. Jake Knapp, John Zeratsky, and Braden Kowitz, *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days* (New York, NY: Simon & Schuster, 2016).
28. See <https://www.ims.gov/grants/awarded/lg-88-17-0171-17>.

29. See www.web3d.org/x3d/what-x3d.

30. See <http://aptrust.org>.

31. See <https://metaarchive.org>.

32. At the time of Chicago Forum, DPN had not announced plans to sunset its services. Thus, the review of DPN as an example of national preservation effort held value to the participants. The LIB3DVR roadmap will need to factor in what it means when national preservation efforts like this end and what libraries should consider in repositories for 3D/VR collection preservation.

33. See <http://mukurto.org>.

34. See <https://www.smithsonianmag.com/smithsonian-institution/replica-tingit-killer-whale-hat-spurring-dialogue-about-digitization-180964483/> for an example of such a situation.

35. Michael Weinberg, "3D Scanning: A World Without Copyright," *Shapeways Blog: 3D Printing News and Innovation* (2016), available online at <https://www.shapeways.com/blog/wp-content/uploads/2016/05/white-paper-3d-scanning-world-without-copyright.pdf> [accessed August 20, 2019].

36. See <https://iif.io/community/groups/3d/>.

37. See www.web3d.org/standards/number/19775-1.

38. COLLADA (<https://www.iso.org/standard/59902.html>) is also an ISO standard file format available for 3D visualization of industrial data but was not discussed in this small group.

39. See <https://www.oclc.org/research/publications/2000/lavoie-oais.html>.

40. The "good," "better," "best" levels of support concept comes from discussions among the working groups of CS3DP who are also considering strategies for tiered implementation of 3D data preservation processes.

41. See DCC's Digital Curation Lifecycle, available online at www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf.

42. See the OAIS reference model, available online at <http://www.oais.info>.

43. See Force11's FAIR Data Principles, available online at <https://www.force11.org/group/fairgroup/fairprinciples>.

44. E. Zierau, *OAIS and Distributed Digital Preservation in Practice* (Kyoto, Japan: 14th International Conference on Digital Preservation iPres, 2017), available online at <http://ipres2017.jp/wp-content/uploads/14Eld-Zierau.pdf> [accessed August 20, 2019].

45. Available at: <http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf>

46. The Consultative Committee for Space Data Systems, "OAIS Functional Entities," *Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-B-1* (Washington, DC: CCSDS Secretariat, 2002), 4–1, available online at <https://web.archive.org/web/20060413232525/http://public.ccsds.org/publications/archive/650x0b1.pdf> [accessed August 28, 2019].

47. See Association of European Research Libraries (LIBER), "Implementing FAIR Data Principles: The Role of Libraries" (n.d.), available online at <https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf> [accessed August 20, 2019]; Alastair Dunning, Madeleine de Smaele, and Jasmin Böhmer, "Are the FAIR Data Principles Fair?" *International Journal of Digital Curation* 12, no. 2 (2017): 177–95, <https://doi.org/10.2218/ijdc.v12i2.567>.

48. Dunning, de Smaele, and Böhmer, "Are the FAIR Data Principles Fair?" 183–85.

49. "How Does DPLA Get Content?" "Can I Contribute Content to DPLA?" and "What Is Metadata?" *Digital Public Library of America: Frequently Asked Questions*, available online at <https://dp.la/about/frequently-asked-questions> [accessed August 20, 2019].