Metadata Requirements for 3D Data

Authors: Jon Blundell, Jasmine L. Clark, Katherine E. DeVet, Juliet L. Hardesty

Abstract

The “Metadata Requirements for 3D Data” chapter provides recommendations for **metadata**[i] based on the five-stage **digital asset lifecycle**[ii]. The “Create” section covers some of the principal ways 3D models are created and discusses what metadata can be captured during the creation process. This section looks at not only what metadata could be captured during model creation, but also why capturing that information is important. The “Manage” section covers the metadata needs for organizing, verifying, and providing **access**[iii] to 3D data. Recommendations include grouping files together as much as possible (by 3D object, by collection of objects, and by project) in order to apply organizational metadata that can be used for access and reuse purposes. The Distribution and Publication section discusses the need for a variety of distribution platforms that support the broadly varying metadata needs of different disciplines. Examples include the need for more granular metadata to support reproducibility and privacy in certain fields, as well as concerns around metadata requirements for accessibility for disability more broadly. Though the circulation and access norms for 3D data are still evolving, the Access and Reuse section posits key metadata anticipated to be useful in the **discovery**[iv] and access of 3D data and models for research or reuse. The “Archive” section utilizes PREMIS as a basis for its recommendations. The rapid changes in the tools and platforms that support the creation and utilization of 3D data results in heavier emphasis on metadata that provides context to data that is often no longer supported by the latest technologies. Additional portions of PREMIS that may be of interest to readers are also specified. The chapter ends with an overall table of recommended metadata fields along with future work needed, naming annotation metadata and metadata for accessibility needs as top priorities for standardization and best practice recommendations.

Introduction

As discussed in previous chapters, one of the ongoing struggles of any technology-based media, including 3D data and objects, is how it is stored, catalogued, and accessed for later reuse. The lifespan of 3D data can be greatly enhanced by the development of best practices for gathering and categorizing 3D metadata, including both information currently recognized by libraries and **archives**[v] as well as creation data, termed **paradata**[vi] by the London Charter[1]. As mentioned in Chapter 2*, Best Practices for 3D Data Preservation,* the activity of metadata generation is a key intervention that should occur at many Preservation Intervention Points (PIPs) throughout the digital asset lifecycle. This chapter will make recommendations for metadata needs for 3D data and objects through the lifecycle stages: Create, Manage, Distribution and Publication , Access and Reuse, and Archive. The digital asset lifecycle described here describes the stages used to manage digital files through their digital asset management (DAM) software[2] and is a simplified version of the DCC Curation Lifecycle Model.[3]

[i] See Glossary
[ii] See Glossary
[iii] See Glossary
[iv] See Glossary
[v] See Glossary
[vi] See Glossary

While not all steps include 3D-specific metadata, working within this framework helps to identify where 3D-specific activities and outputs occur and what metadata is needed to record and track within those stages of the lifecycle. The recommendations include the types of metadata needed (names or dates, for example) but do not specify a particular metadata standard to use or controlled vocabularies to apply. This work is based on experience with metadata and 3D object creation and collection management from the CS3DP community. Examples from the community and feedback on the metadata recommendations will also be shared. While examples include a mix of metadata standards used for different types of information, future work is needed to gather consensus around standard metadata properties and controlled vocabularies for 3D models and collection management. Within the Good / Better / Best framework of CS3DP recommendations, this chapter offers something between Good and Better, with Best being common metadata standards adopted and used for all 3D object creation and collection management.

Methods

The recommendations made in this chapter are based upon data collected via two surveys along with feedback solicited via the second Community Standards for 3D Data Preservation (CS3DP) forum at the University of Michigan. The first survey asked for stakeholders working with 3D data for their current metadata practices and models. These responses were then collated into a Google Sheets spreadsheet and categorized by the types of data they described (Project, Model/Data, Pre-Processing/ Processing, Capture,[vii] and Original Item). These results were then shared at the second CS3DP forum. To aid in our understanding of the information needed by different categories of stakeholders, we collected user profiles and user stories, much like the case studies seen in Chapter 2, *Best Practices for 3D Preservation*, taking care to consider individuals who may interact with 3D data at all points within the lifecycle of that data.

Attendees' of the second CS3DP forum brought a wide variety of perspectives on the creation and management of 3D data based on their role in the process (such as creators, publishers, or **repository manager**[viii]s) and the **workflows**[ix] and technical lexicon of their particular community or institution. These diverse frames of reference led to confusion regarding the initial survey results among forum attendees when the survey's terminology and assumed workflows did not align with the forum attendees' practices. Based on this feedback, the decision was made to organize this chapter and its related metadata recommendations according to the digital asset lifecycle. This was intended to ease data collection in our second survey, as respondents could focus on their respective areas of expertise, while also making the results easier for readers to navigate. The second survey gathered data regarding the metadata fields collected during the first survey, while also acknowledging potential gaps and asked respondents to identify additional metadata fields they thought were missing from the survey.

Considerations, Decisions, and Scope

Responses to the surveys were limited in number and disciplinary representation was equally limited. There were ten responses to the first survey with anthropology, archaeology, geology, and museums being the represented disciplines. There was a lack of data regarding utilized schemas, tools, phases of capture, workflows, and objects designed and modeled digitally (not scans of physical items).

[vii] See Considerations, Decisions and Scope for more information.
[viii] See Glossary
[ix] See Glossary

We received eight responses to the second survey and will discuss the results in each section of this chapter. There is also a need to acknowledge that the community of stakeholders producing large quantities of 3D data is limited in size and still evolving. The data captured within this chapter should be considered a snapshot of current 3D practices in the academic and cultural sectors.

Consideration was given to different methods of capture and creation along the lines of those mentioned in the Chapter 2, *Best Practices for 3D Data Preservation*, but in attempting to list and address various methods, it was found that extensive knowledge of common/assumed workflows and technical requirements for each method would need to be articulated to prescribe metadata recommendations for each method (i.e. the metadata needs for CT data, as opposed to photogrammetric data, are distinctly different). This was determined to be out of scope for this chapter and, as a result, this chapter will focus on examining the creation metadata elements common across all production methods.[x] The information available in this chapter pertains to commonalities and unique fields related to the general categories of reality capture, sources based, and artistic 3D data. While the "creation metadata" discussed in this chapter could be viewed as what the London Charter refers to as "paradata," for the purpose of these recommendations, this information is considered to be another form of metadata that is able to be captured and catalogued.[4]

Digital Asset Lifecycle and 3D Metadata

Create

Metadata in the Create stage is associated with the process of collecting/capturing source data and the process of model construction (process inputs), the finished 3D model (process outputs), and, for models that are digitized versions of physical items, metadata associated with those physical items. Metadata capture needs to begin as soon as a project involving creation of 3D models (or 3D data in general) is conceptualized. This is because the project's intent and how the resulting models will be used[xi] strongly inform the type and level of documentation required. Central to understanding the metadata needs of a given 3D model is knowing the method used to create it. Creation documentation needs for a reality capture model, created from 3D scans or **photogrammetry**[xii], will significantly differ from a sources-based model, created manually using reference material. A purely artistic model's documentation needs vary further still. A model can be created using any mix of these approaches and there are other approaches, such as procedural modeling, that, while outside of what is discussed in this chapter, would still need to be considered if applicable. Of those surveyed, every respondent documents the project for which a model was created, as well as its method of creation/acquisition, [xiii] a practice which this chapter recommends.

Stakeholders

Model creation documentation is primarily the concern of those producing 3D models and, for models that represent real world objects and environments, the holding entities and stewards of those subjects. Once created, this documentation is most relevant to those with expertise in digital 3D

[x] For examples of more granular, method specific metadata, see responses to the first survey in this chapter's appendix.
[xi] CS3DP Second Metadata Survey: Overview information - Objective/Purpose 50%
[xii] See Glossary
[xiii] CS3DP Second Metadata Survey: Method Used - Method of capture/acquisition 100%; Project identifying information - Title 100%

technologies and those who need a deep level of understanding of a 3D model such as those with research, academic, or other technical interests in the material.

Survey

As indicated by the survey respondents, there are myriad ways to create 3D data. Each process requires specific approaches to capturing metadata that will be useful for the data's **preservation**[xiv] and (re)use. These 3D data creation and capture methodologies are outlined below, as well as possible variations of metadata one should document as part of these processes.

General Creation Metadata

This chapter focuses on the three categories listed above; reality capture, sources based, and artistic. While each has specific metadata needs, there are common elements that are seen across all three: the basic who, what, when, and how (with "where" depending on the subject of digitization). Of those surveyed, all reported recording a name/description/identifier of what was created, actors involved, tools/software used, and the dates when actions happened.[xv] These are not 3D specific terms, but broad terms that can describe most any act of creation, thus there is no need to use new or exotic metadata elements to capture this information. Commonly-used **descriptive metadata**[xvi] standards such as Dublin Core, Darwin Core, VRA Core, or MODS will cover this information, though a possible expansion of controlled vocabularies could be needed depending on the standard used. It should also be noted here that, if there is a real world object or environment that a model is derived from, the necessity of metadata to describe that real world object or environment will depend on whether or not the real world object is described in a digital system elsewhere. If real world description is needed, one of the commonly-used descriptive metadata standards mentioned above, or a relevant domain specific metadata standard should be used. Depending on the data management approach the system which stores 3D data and models may not need to store information about real world counterparts if it exists in another system of record, other than a way to identify that real world thing (preferably using a resolvable globally unique identifier). This idea is supported by the survey, where all respondents identified the real world object or environment[xvii] a model was based on with a unique identifier, when applicable.

Metadata elements that track the finer-grained steps and decisions that go into a given creation (processing) action[xviii] are used less than the basic who, what, when, and how metadata elements, as reported by the survey respondents. This is reasonable as there are two major logistical challenges to this higher level of documentation. First, and maybe obviously, documentation becomes more time consuming and burdensome as reporting requirements increase, decreasing the likelihood that the documentation will be undertaken. Second, the more granular creation metadata is, the more tasks and tool specific it needs to be. An effective metadata reporting strategy and schema would go a long way towards facilitating more in depth creation documentation, of which there are few, if any, accepted community standards.

[xiv] See Glossary

[xv] CS3DP Second Metadata Survey: Model Identifying Information - Title 100%, Unique identifier 87%; Linked Fields - actor 71%; Resources - Processing software 85%; Processing Action - Date of action 100%

[xvi] See Glossary

[xvii] CS3DP Second Metadata Survey: Original Item - Unique identifier 100%

[xviii] CS3DP Second Metadata Survey: Processing Actions - Processing decisions 57%, Action Method 42%, Action description 57%, Action 14%

Creation Transparency and Reuse

While acknowledging the burden of process documentation, such documentation is crucial for any models used in a published work or a scientific/academic setting. The source information for a model, as well as the manipulations and interpretations of the source data that contributes to a model's creation, is critical to understanding what a given model represents, and importantly, what it does not. Both reality capture and source based models originate from collected information, whether that information is scan data or reference observations and media. Understanding how this information is interpreted and manipulated to create a 3D model is key to understanding what elements of that model are representative of the source information and what elements are artifacts of processing/interpretation. As part of this, it is essential to recognize that any reality capture or source based model is modified by the lens of interpretation, whether that interpretation is coming from an algorithm in a piece of hardware/software or from a human making judgement calls while interacting with that hardware/software. The goal of tracking the steps that go into a model's creation is to give transparency to these interpretations, allowing those that interact with the produced model (for comparison, illustration, investigation, etc.) to fairly assess the model's trustworthiness for their needs.

Thus, it follows that any 3D models created by reality capture or modeled from sources which purports to represent a real world object or scene should have as much creation documentation as is reasonable. Capturing creation metadata also allows for repeatability of the creation process, so the same or similar enough model may be derived using the same source data and same processing methods. There is a balance that must be struck between data transparency and workflow efficiency when creating and presenting 3D models as digital facsimiles of the physical world. How to arrive at that middle ground is something that creators and repositories must currently decide for themselves, and should be a focus of community standards development in order to better facilitate the sharing and reuse of resources and data.

Metadata for Reality Capture Models

While reality capture and manually created source-based models share many core metadata features, each has its own needs and considerations. The source information for reality capture models is the raw data off of the capture device (laser scanner, CT scanner, camera, etc.). This raw data could be in the form of **meshes**[xix], **point clouds**[xx], and/or images. It could be output in open, accessible file types or closed proprietary file types. While the files that are produced from a scanning device may be termed raw data, it should be recognized that they have undergone some type of interpretation on the capture device to convert the electrical sensor readings to an intelligible format such as those listed above. While some of this interpretation is opaque to the end user, whether because it is a trade secret or simply not provided by the tool, capture equipment should have device parameters or user definable settings which can be documented. As many tools and software do not encode parameters and settings used into the files they output in a user accessible way, it should be standard practice to document any applicable parameters and settings, as well as basic information such as the make and model of the equipment used, as part of recording capture metadata.[xxi] Along with documenting capture tools and their settings, it is also helpful, depending on the capture technique, to document the capture process and the intended strategy for data processing. Examples include, but are not limited to, if targets were used for aligning multiple captures types such as laser and photogrammetry, if a spherical or focal

[xix] See Glossary
[xx] See Glossary
[xxi] An example of a tool that does include parameters and settings in its output files would be a digital camera via embedded EXIF metadata.

stack rig was employed, if color calibration targets were used, or if there are calibration or rig scaling datasets.

Metadata for Manually Created Source-based Models

The reference material for source-based, manually created models could come from a variety of resources. Examples could include measurements of existing or partial artifacts, sketches of hypothesized historical representations, written documentation from research sites, and photographs or other media sources depicting the object or environment (or even similar/related objects and environments). How these references are used and interpreted needs to be captured and cited, just as it would in academic writing. As with reality capture models, it is critical to understand what sources were used during the creation of a model. For example, this information would allow someone viewing a historic reconstruction to understand what parts of the model are based on documentation, and what elements have been filled in from imagination or interpolated from relevant sources to complete the reconstruction. Citation possibilities are discussed later in this chapter in the Access and Reuse section.

Model Processing Metadata

Once capture and/or source data documentation is complete, processing documentation needs to be considered. Of the survey respondents, roughly half reported recording processing decisions, action methods, and descriptions.[xxii] This is understandable as the approaches and software that can be used to process a model are vast and varied. Also the underlying algorithms and their inputs which are used to manipulate 3D data are opaque in most commercial software (an exception is many photogrammetry software packages allow for detailed reports to be generated which include settings and other metrics).[xxiii] This makes process documentations complex and difficult to standardize. Documenting software and source datasets, as well as any large decisions/modifications such as noise reduction and hole filling will go a long way towards model transparency. Citing a followed best practice, such as the BLM and CHI photogrammetry workflow, can also provide significant insight into the processing workflow that created a model.[5] Another facet of processing documentation is if down sampled models are needed to facilitate use, such as derivative models made for web viewing or 3D printing. In this case, it is important to be able to document the source model for any derivative models created, as well as the methods of derivative model creation, so this information can be passed on to the model's consumers.

Metadata for Artistic Models

For artistic creations, the creation documentation needs will vary widely depending on the use case and the creator's intent. As works of art and artistic representations could be intended for any number of modes of consumption, both digital and physical, the information needed to codify the digital representation and associated files could vary widely. For the most part, creation/capture data will be similar to that needed for capture-based or source-based models, depending on the creation modes and elements used. These models may also be a part of a time-based media experience, gallery or artistic installations, interactive experiences, and games. Taking this into account, relevant exhibition or other contextual information may also be necessary to note.

[xxii] CS3DP Second Metadata Survey: Processing actions - Processing Decisions 57.1 %, Action Method 42.9 %, Action Description 57.1 %

[xxiii] It should be noted that some 3D processing software can generate process and quality reports, and if available these reports should be stored with the respective model.

Creation Metadata Example

A common 3D model creation method is digital photogrammetry. The source information for this creation method is a collection of digital images. Examples of creation metadata related to this source information would be:

- the subject of digitization (e.g. descriptive information on the subject of capture, identifiers pertaining to the subject)
- equipment used (e.g., camera make and model, scale bar measurements), equipment setting (e.g., ISO speed, f-stop)
- capture approach (e.g., was a camera array used, was cross polarization used),
- and post processing actions performed on the images before they are brought into a photogrammetry software package (e.g., was chromatic aberration correction applied, was image sharpening applied)

Depending on the file types used, some of this information might be stored directly in the image files (e.g., EXIF metadata), some might be stored in standard digital photography **sidecar**[xxiv] files (e.g., XMP files), and some information will have to be recorded outside of standard digital photography tools and conventions.

Examples of creation metadata related to the model processing would include:

- which images were used during model creation
- how these images were used (e.g. was an image used during alignment, surface reconstruction, and/or texture mapping)
- what processing actions were taken (e.g. camera alignment, geometry smoothing) along with the software used and input parameters for those actions
- and qualitative information characterizing the process (e.g. the project's average reprojection/alignment error)

While some software will produce limited reports on the model creation and editing processes, for the most part this process information does not have a standardized way of being exported from processing software or generally being documented. Much of the creation metadata will have to be recorded in an outside tool (e.g., spreadsheet, MorphoSource, the Digital Lab Notebook) in either a structured or narrative format.

At this point, the process creation metadata has been addressed. The remaining metadata of interest as it relates to a produced 3D model is mostly technical information (e.g. model size, model **resolution**[xxv], UV map types) as well as information about the subject the model represents. Without delving into the specific technical aspects of different 3D file formats, it is worth noting some formats lack the ability to fully describe themselves in relation to types of information they may contain as well as the numerous methods by which a 3D model can be rendered. Because of this, it might be necessary to store additional **technical metadata**[xxvi] outside of the model files. For example, many 3D model file formats store **texture maps**[xxvii] in separate individual image files and not all 3D file formats fully describe all possible texture map types. Thus, it might be necessary to record how a texture map should be used (e.g. diffuse color, normals, ambient occlusion) separately from the model files. It is also worth mentioning that few if any 3D model file formats store scale information (e.g. are the models units in millimeters, feet, etc.) so this metadata would also need to be stored separately.

[xxiv] See Glossary
[xxv] See Glossary
[xxvi] See Glossary
[xxvii] See Glossary

Implications

As with most metadata collections, the bottom line is that data creators need to take the time to record the relevant information in their workflow before it is 'lost'. Creators should look to leverage tools to make their metadata collection tasks easier, whether that is creating 'shoot sheets' or using a tracking program such as CHI's digital lab notebook.[6] Much of the burden of creation documentation could be eased if tools and software self-reported by providing action logs in a non-proprietary format, though at the time of writing this is largely not the case. Lastly there needs to be realistic expectations on how much information can be reasonably tracked. One survey respondent reported the SHAPES project [Sharing and Helping Academics Prepare for Educational Success], housed out of the Texas Tech University Libraries, started with over 70 metadata fields describing catalogued 3D models but after realizing the massive task of reporting to this level, the schema was reworked to a more manageable 18 fields.[7] In their words, when deciding on an appropriate schema "find the fields that are the most important and focus on those, otherwise it'll take someone an hour to do one record, and that kind of time in the metadata processing denotes an inefficient schema." While there is no prescribed amount of time necessary to create metadata, the point here is that context matters and considerations should be made based on expected use of the 3D model and available resources for metadata creation. specifically what metadata is needed to facilitate the expected downstream use of the 3D data and models as well as who is doing the work of metadata creation, how much time they have available for the task, how much work it will take, and what tools, if any, are available to ease the work.

Recommendations

The recommendations below are separated based on the level of effort required and rigor of intended reuse. Recommendations for good practice assume basic access and use of datasets by casual users, addressing discovery, context, and citation needs. Recommendations for better practice facilitate informed use of datasets and 3D models in a research or academic context where judgments on data quality/suitability are required. Recommendations for best practice are aimed at full reproducibility and might require a high level of technical experience.

Good

Document and use a folder structure and file naming convention to organize source data and 3D models for management purposes. For example, group data and models by digitization project, collection, etc.

Include the following information as structured data in README.txt files, CSV files, spreadsheet, etc.

Project level information:

- *Project name*
- *Project identifier*
- *Project date*
- *Description/abstract*
- *Project authors/creators*
- *Stakeholders/contributors*
- *Project rights information*

Source data information, if a reality capture model or sources-based model:

- *Method of creation*
- *Creation date*
- *Information identifying real world object or environment*

- *Creators*

3D model information:

- *Subjects*
- *Creator*
- *Geometry information (e.g., number of faces, bounding box size, etc.)*
- *Textures (what types of UV maps are available)*
- *Materials (If any material properties are applied to a model)*

Better

Include the following information as structured data in README.txt files, CSV files, spreadsheet, etc.

Source information for reality capture or source based models

- *Resolvable GUID for records for subject or sources*
- *Data sources*
- *Georeference information if applicable*

Source data creation information

- *Capture device make and model*
- *Capture event details*

3D model processing information

- *Source data used*
- *Software used*

Documentation of capture and processing workflows

Best

Include the following information as structured data in README.txt files, CSV files, spreadsheet, etc

3D model processing information

- *Detailed steps and log outputs*

Use standardized metadata properties to define project, file, and 3D model properties listed above

Manage

The Manage stage within the digital 3D asset lifecycle is meant to ensure that those charged with maintaining 3D models and data have access to those files and accompanying metadata. Metadata in the Manage stage is associated with reviewing, annotating, and approving activities along with version control and the logistics of giving people access to view or create annotations on a digital asset. Digital asset management system software often provides roles with appropriate access levels that can be assigned to individuals or groups and can be useful for this purpose if other needs for storing and accessing 3D models are also met. Metadata at this stage is often associated with reviewing files to verify that what was expected has been received. Creating **checksums**[xxviii] (algorithmically generated sequences of numbers and letters that represent the data contents of files) before files are passed along to this stage can make this type of verification a fairly quick and objective task.[xxix] Received files can be verified using a checksum calculator to compare and ensure that nothing has changed about the received files during transmission. Metadata also helps at this stage for activities such as: determining intellectual property rights to establish how a 3D model can be accessed and reused; annotating 3D models with descriptive metadata; approving activities to verify any modifications; version control; and logistics of giving people access to view and comment on a digital

[xxviii] See Glossary

[xxix] See Glossary for further definition and explanation of checksums.

asset. Allowing others such as curators, catalogers, information specialists, or subject matter experts to access digital assets at this point provides the ability to enrich the 3D model, add internal cataloging notes, or cite external resources.

Stakeholders

The main stakeholders with management concerns are collection owners and collection managers who will be working directly with digital objects to review, annotate, and approve that content is stored and available as expected. These stakeholders will also determine the appropriate level of access to set for public availability - what format(s) will be available and how widely.

Collection owners are likely to be in charge of multiple collections and can have a variety of concerns for managing collections online. Examples of collection owners include archivists, museum curators, archaeologists managing an excavation site, artists, game developers, or biologists with research specimens. Collection owners will be particularly concerned that access levels are set appropriately for 3D models and their accompanying files and metadata (if everything should be fully available or if restrictions should be applied). They will also want to ensure that all 3D models are present in the system, that they accurately represent any physical objects on which they might be based (for example, all files are present when compared to a manifest with checksums), that 3D models that are not reality-based are completely represented (for example, all 3D models that should be part of a Virtual Reality environment are included), and that everything is described accurately and appropriately.

Collection managers will not necessarily be stewards of a collection of 3D models but could be in charge of managing online access across various collections within a single digital asset management system. Examples of collection managers include digital collection service managers (system administrators or those who manage a **digital repository**[xxx] or digital asset management system used by several collection owners), web developers, librarians, or others involved in information technology management. Collection managers will be concerned with ensuring all necessary files and required metadata are present and that a 3D model or collection of 3D models has enough information to be discoverable within a larger system or set of models and collections.

Survey

The initial use case survey identified metadata properties associated with managing 3D data. Based on this information, the follow-up survey asked questions about specific properties that can help identify a project or collection associated with a 3D object and provide high level information (an overview) of the project or collection, identify people with various roles (such as creator, contributor, stakeholder), and supply copyright information that can help establish access rights for a project or collection of 3D objects.

Implications

The metadata to establish management capabilities needs to identify the overall project for discovery purposes. This allows 3D models or sets to be grouped and organized for discoverability by project and, if possible, permissions set based on applicable rules (i.e., copyright, embargo, etc.) if they can be applied to these same sets as a whole. This in turn helps establish levels of access for an entire group of 3D models which then helps establish access to individual 3D models.

[xxx] See Glossary

Determining copyright and licensing inherent in an object is important to properly managing a 3D model or collection of models. The information provided in Chapter 5, *Copyright and Legal Issues Surrounding 3D Data*, addressing rights and licensing should be referenced to help make appropriate determinations regarding what content is copyrightable and how 3D models and their associated data can be accessed and reused. Determining that a 3D model is protected under copyright may mean the files and creation information that make up that 3D model are not available for duplication and reuse but derivatives or portions of the model might still be viewable. Additionally, metadata describing the 3D model can still be used for discovery and aggregation purposes, making the 3D model discoverable even if it is not available for viewing or download.

Inherent in providing access to different files is knowing something about those files or about the data as a whole to understand what it is and how best to supply access. This requires, not only appropriate capture metadata, but effective validation workflows to verify the datasets are complete and uncorrupted. Validation is also necessary to ensure that appropriate technical requirements are in place to ensure access to end users. For example, original scans or point cloud files can be large enough that online delivery is not feasible even if the rights determination is that they should be openly available. Even when using a digital repository or digital asset management system, providing different levels of access for the different files involved in a single 3D model can prove complicated.

Tools like Meshlab and Blender are available for evaluating technical aspects particular to 3D models such as number of faces, number of points, number of slices, and spacing between slices. Digital repository systems such as MorphoSource from Duke University or the systems used to manage collections at the Smithsonian Institute have used tools such as these as part of their workflows, but there are no set minimum standards for 3D data preservation yet. These tools have only been used so far during upload/ingest processes and not as part of verification for a 3D model or an entire set or collection of 3D models. This means the effectiveness of using these tools to verify technical information at scale for multiple models or an entire collection of 3D models as a preservation activity is unknown. Scripted options for evaluating batches of 3D models are possible (working with a command line tool such as meshlabserver, for example) but a minimum standard for 3D models is needed for the Manage stage of the digital asset lifecycle so that ingested models can be approved and accessed. Establishing internal minimum standards and using tools that can extract this kind of metadata will help to organize a collection of 3D models and determine how best to provide access to an individual 3D model.

Recommendations

The survey results show that there is some preference for certain types of project level information and fields: project name; project identifier; project date, description/abstract, file/data types, and subjects; project authors/creators and stakeholders/contributors; and project copyright information. In the area of management, the fields identified are not particular or specialized for 3D models or data. This is a benefit to managing 3D collections in that the types of information needed and activities performed are similar to other digital collection management needs and processes and might be compatible with available content management system software. Because managing access by users or by groups to an entire collection or data set is generally a necessary part of managing any type of digital collection, this sort of functionality is already provided in many content management systems, such as institutional repositories or digital repository systems.

If 3D models can be managed as sets or collections, copyright management could occur at the set or collection level, applying copyright status that then indicates appropriate access levels to set for a grouping of 3D models. This might not always be effective if individual 3D models have rights maintained separately or have rights connected to a physical object used to create a 3D model, so it is worth considering if groups of 3D models can have the same copyright applied and if the management software includes that capability. If different resolution derivatives of 3D models require different access levels (i.e., an access level derivative is openly available for viewing but the files and creation information that make up that derivative are restricted to only authorized users), that might require a different management model that can store the openly accessible 3D model with information that provides a pointer to a more restricted location for the files and creation information behind it. Please reference Chapter 5, *Copyright and Legal Issues Surrounding 3D Data*, for a complete discussion of concepts concerning rights for a complete 3D digital object versus rights for the parts that make up that object. Ensure that copyright statements are understandable to end users accessing these objects and in any place where these models are shared or aggregated outside of the content management system by using standardized rights (such as those at rightsstatements.org) and licensing information (such as Creative Commons).

Managing objects in a digital collection involves verifying that what was expected was received. This often includes checking technical metadata about the files themselves (i.e., filename, checksum, file format). The Library of Congress offers recommendations for metadata standards and protocols to use (although 3D is not included at this time) and annually updates a Recommended Formats Statement that should include 3D soon.[8] Open source communities working on digital collection management also offer good resources. The Samvera Community offers recommendations for baseline technical metadata that should accompany any digital media file, including 3D models.[9] The Smithsonian Institute offers a 3D metadata model defining technical metadata fields in use there, as does the Archaeology Data Service / Digital Antiquity Guides to Good Practice.[10] Different types of digital objects will also have technical information specific to that object type that can be used to check the item being received. Although no preservation standards exist at this time, 3D models have similar technical features that can be used to verify that the minimum standards are met for that type of digital object (such as scale, number of faces, number of points, number of slices, spacing between slices).

Based on survey responses and additional research, recommendations for “Good” reflect file organization that does not require a digital asset management system and supplying accompanying metadata in associated text files. Recommendations for “Better” reflect use of a digital asset management system that supports user role definition and access features along with a way to define projects for grouping models and storing metadata about those models. Recommendations for “Best” reflect implementing standards for associated metadata when possible.

Good
Group 3D models together as project or collection for access and management purposes
Include following project level data:

- *Project name*
- *Project identifier*
- *Project date*
- *Description/abstract*
- *File/data types*
- *Subjects*
- *Project authors/creators*

- *Stakeholders/contributors*
- *Project copyright information*

Supply copyright information at highest level possible (project/collection/object/file).
Using available tools,[xxxi] *document the following properties from files in a structured file type such as a README.txt file:*

- *File-specific:*
  - *Filename*
  - *Checksum*
  - *File format*
  - *File size*
  - *File creation date*
  - *File modified date*
  - *File version*
- *3D model-specific:*
  - *Scale*
  - *Number of vertices*
  - *Geometry type*[xxxii]
  - *Number of faces*
  - *Number of points*
- *CT data-specific:*
  - *Number of slices*
  - *Spacing between slices*

Better
Use digital repository/digital asset management system with access capabilities to define permissions and make project/collection discoverable.
Store project, file, and 3D model properties listed above with 3D models in digital repository/digital asset management system.

Best
Use rightsstatements.org statements and CreativeCommons licenses to define standardized access and reuse levels.
Use standardized metadata properties to define project, file, and 3D model properties listed above.

Distribution & Publication

The Distribution and Publication stage encompasses the licensing, sharing, and dissemination, of 3D data. The metadata associated with this stage are essential for discovery and, to some degree, will depend on the intended use. It should be human readable and machine actionable so that data can be identified and located via a **persistent identifier**[xxxiii]. Recommendations are driven by distributor type, often related to the user profile (e.g., casual, academic, etc.).

[xxxi] Example of available tools include DROID (https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/) and FFProbe (https://ffmpeg.org/ffprobe.html) but there are others available to also do this work.
[xxxii] For example, polygonal quads, polygonal tris, polygonal ngons.
[xxxiii] See Glossary

Stakeholders

Distribution metadata is first and foremost aimed at end users. As a result, publishers, websites/ **aggregators**[xxxiv], and repositories are all stakeholders as they attempt to support access and re-use. Additional stakeholders include creators looking to track the reuse of their work, those looking to determine the trustworthiness and source of 3D data, and those looking to identify additional work from relevant creators. Additionally, there are a variety of industries in which 3D data is used, and a variety of uses to which it can be applied. The Science, Technology, Engineering, and Mathematics fields and related industries (medical, automotive, meteorological, etc.), as well as the Arts and Humanities (anthropology, art history, psychology, etc.) are among the many disciplines and professions utilizing 3D data. Each of these stakeholders has differing distribution and metadata requirements reflective of the norms and expectations around reuse within their particular discipline or profession. In response to this reality, there are variations in existing commercial, academic, professional, and recreational distribution platforms, to support different types of metadata. However, there is plenty of room for further development, especially as 3D data formats, and their uses, continue to evolve.

Survey

Responses to our first survey, along with ensuing research, indicated that many of the metadata requirements for the distribution (or publication) phase of the 3D life cycle are not 3D specific. Metadata supporting fair use and discoverability, including copyright information and unique identifiers, is already fairly standard for other digital formats like digitized photos and documents.[xxxv] Our second survey looked to determine if any of this standard metadata was not applicable to stakeholders, while also providing the opportunity to share additional metadata that was not already represented in the survey. We asked stakeholders whether they used, or would use, DOI/PID/Location, Copyright, Embargo, and/ or Citation Information. None were listed as undesired. There were also no additional fields provided. However, it must be considered that our survey responses retain galleries', libraries', archives', and museums' biases.

Implications

As previously stated, distribution of 3D data requires metadata that supports copyright, controlled access, and the measurement of impact. This typically takes the form of unique identifiers, embargoes, and copyright statements, all commonly used in other forms of publishing. However, some additional needs that should be considered are distribution derivative and accessibility metadata. Derivatives for the purpose of general distribution should have metadata generated and applied at the point of creation. Whether generated by a digital asset management system or created manually, the derivative should have the same information recommended in this chapter for the originating dataset. However, the source information would point to the originating dataset and the originating collection, item, or material the derivative represents in order to provide context to the derivative's creation and purpose. Accessibility metadata should include fields for alt text, long descriptions, and information on the location of, or means of requesting, accessible alternative formats. This poses a challenge in

---

[xxxiv] See Glossary

[xxxv] Dublin Core (https://www.dublincore.org/specifications/dublin-core/), MODS (http://www.loc.gov/standards/mods/mods-overview.html), RDA (https://www.rdatoolkit.org/about) and other schema have all been used with MARC (https://www.loc.gov/marc/) and other encoding standards to create metadata records for both digitized and born-digital audio, video, photographs, etc. Aggregators like DPLA (https://dp.la/) and Europeana (https://www.europeana.eu/portal/en) have multiple examples of these types of records from various organizations.

systems that do not support these features and should be considered by stakeholders when deciding between creation, hosting, distribution, and preservation platforms, as well as metadata schema.

When it comes to distribution platforms that support both the viewing of content, along with the more complex metadata demands of certain disciplines and professions, affordable options are limited. Distribution demands for casual, non-professional use will be different from the demands of industries and disciplines where dataset **accuracy**[xxxvi] can be the determining factor in safety and crucial decision making. 3D models, virtual reconstructions, aerial photogrammetric scans, and CAT scans are among the various formats that can be shared in digital repositories, commercial platforms, journals, and other distribution platforms. Each of these formats and their associated function will differ in the ways they handle copyright, access control, and tracking due to different legal requirements and desired reuse.[xxxvii] However, while distribution metadata needs will vary from sector to sector, the limited availability of affordable hosting means that available platforms, like Sketchfab, often have a mixture of model types, yet limited metadata options, complicating metadata application and retention for data with more robust metadata needs. There are more specialized distribution platforms like the *NIH 3D Print Exchange*, which specializes in hosting medical 3D printing models for download, but these are meant to accommodate a very specific subset of 3D data for a very specific subset of use. Datasets utilized for academic, cultural heritage, or medical purposes that do not have printing as their sole purpose, or at the core of their use will require very specific display, manipulation, and granular metadata support.

Recommendations

The recommendations in this section are based on the intended use of the dataset in question. Recommendations for good practice will be based upon reuse of datasets for recreational virtual reality, 3D printing, or other casual, informal purposes. Recommendations for better practice will be based upon datasets designed to support research by serving an illustrative or reference function, like historic recreations. Recommendations for best practice will be based on datasets designed to be used as research data for fields and industries that are dependent upon accurate standards for reuse such as meteorological and safety simulations, and CAT scans in the medical field. All metadata from the previous sections should be carried into this one.

Good

For casual use, users need to understand the format and intended use of the model. Intended use will usually dictate whether a model is suitable for printing, virtual reality, or other purposes. For an example, see Sketchfab.[11]

-System Unique Identifier
-Author(s)/Creator(s)
-File type
-File size
-Geometry information (quads, triangles, etc.)
-Vertices
-Textures/Materials
-Description
Accessibility information
    -Alt Text
    -Long description

[xxxvi] See Glossary
[xxxvii] See Chapter 5 *Copyright and Legal Issues Surrounding 3D Data* for more information on this topic.

-Accessible format/information location (alternative accessible format or accessibility policy information)

Better
3D data meant to act as an illustration in scholarship should include the previous data in addition to the recommended fields below. For an example see the NIH 3D Print Exchange.[12]

-Link to original ***source material***[xxxviii] *OR formal name and information of original item (scientific name, formal name, etc.)*
-All processing done (textures added, artistic liberties taken)
-DOI/PID
-Associated works
-Citation information

Best
Datasets designed to be used as research data for fields and industries that are dependent upon accurate standards for reuse should include the previous data in addition to the following fields. For an example, see Morphosource.[13]

-Detailed Capture Metadata (amperage, axis resolution, scanner resolutions, exposure time, wattage, etc.)
-Project information:
  -Funding information
  -Affiliated organizations
-Technician information
-Checksum

Access & Reuse

Metadata in the Access and Reuse stage is associated with the circulation, consumption, citation, and use of existing 3D data by end-users and non-practitioners. The purpose of these data are two-fold. In one case, these data are to facilitate ease of access, (re)use, and interoperability (ideally) of systems once the data are identified. These data also allow for appropriate citation, underscoring the integrity, authenticity, and **provenance**[xxxix] of the 3D data. A user should be able to decipher, from these metadata, whether the 3D data is of use to them. The Retrieve stage of the digital asset lifecycle is key in the continued vitality of any information source, including 3D data.

Stakeholders

Stakeholders in the access and reuse process come from all stages of the lifecycle and have vastly different priorities when it comes to accessing and consuming 3D data. First and foremost, content creators of all disciplines--the sciences, cultural heritage preservation, the arts, or fields that are yet to be developed, are invested in how they can facilitate the reuse of their data, regardless of how it was created. Both to preserve research and for the benefit of those consuming data later, it is necessary to create metadata recording the creation of the work as mentioned earlier in this chapter in

[xxxviii] See Glossary
[xxxix] See Glossary

a way that is both archival and understandable. This metadata should be filtered and keyworded in ways that aid in the discoverability of a given dataset. The academic community is still formulating the research norms and preservation standards for 3D data using foundations such as the FAIR (findable, accessible, interoperable, reusable) Data Principles[14]. While it may not seem immediately useful, given research patterns of other data, it is conceivable that researchers may find it useful to also be able to search larger groups of data by creation mode, source material(s), related derivatives or object size/scale in order to better serve their research. Similarly, holding entities of 3D data take a vested interest in the accessibility and discoverability of the items and data within their catalogs, in order to better facilitate the access and reuse of that information.

On the path to publication, researchers in higher education and industry need reliable discoverability to find 3D data and models, verifiable provenance and tools for recreation of both data and models to ensure the validity and accuracy of reproducible data. This may include a sort of intellectual chain of custody listing the researchers that have worked on an object and the methods used as well as recommended software or viewers to access the item. In addition to this, the academic community will need a means to cite the 3D data being utilized as no sufficient standards for doing so currently exist, regardless of whether that data is open access or rights restricted. As these researchers may or may not be 3D practitioners themselves, consideration will need to be given to common language, dataset types, and creation formats,as well as standardized citation formats. In addition to general standardized language, this consideration should extend to discoverability for information specialists and research assistants aiding those researchers. Information specialists and librarians who are not 3D specialists will need to be able to assist patrons of all types in finding data through common websearch, database searches, as well as integrated library systems [ILS] and discovery layers provided by a variety of vendors.

Beyond creators and researchers, a wide variety of consumers within the general public have interest in accessing 3D data. Some may be artists or enthusiasts, looking to use 3D data for personal enrichment or to augment personal projects. Teachers in primary and secondary education may look to 3D data to augment concepts taught through classroom instruction. Students in those classes may look to it to aid in class projects or to fuel further study in personal interests. Furthermore, with the advent of easily accessible resources such as Thingiverse, MorphoSource and the Smithsonian Digitization Program Office, and the increasing access to 3D printers, individuals from various backgrounds may have an interest in discovering 3D models for self-guided learning and 3D printing. In general, these consumers are less concerned about the ability to recreate the final project than the ease of finding the item using natural language or keyword searching.[15]

Survey

As universal standards involving the access, use, and resource sharing of 3D data and models has not yet been codified, survey responses regarding access of said items was limited to a record of downloads and use count as well as, potentially, fields to record who may have edited or processed the available files. While programs like SHAPES [Sharing and Helping Academics Prepare for Educational Success], housed out of the Texas Tech University Libraries, are working to establish resource sharing standards, adoption is slow. There has been more progress in open access resources like MorphoSource and Thingiverse but access metadata for these databases is also currently limited to download counts.[xl] As resource sharing of 3D data and models evolves, it is likely the related

[xl] While MorphoSource and Thingiverse have different funding origins--Morphosource is housed at Duke University and Thingiverse is run by corporate entity MakerBot--both repositories serve as sources of 3D data and models that are readily and freely available to the general public.

metadata needs will as well. With this in mind, current recommendations for Access and Reuse will focus on the metadata needed to locate and access a given 3D dataset or model, rather than that generated during the access process. As community standards for the access and sharing of these resources develop, necessary metadata beyond page views and download numbers may develop as well.

Implications

Data that cannot be found cannot be used and, therefore, has no longevity. In order to continue to utilize 3D data, it needs to be discoverable and accessible by both the academic community and the general public. As resource sharing for these formats is still developing, it remains to be seen what sort of circulation and access data should be preserved but the topic should definitely be considered as the prevalence of 3D data increases. Though 3D data creation is far from new, access processes for 3D data are still developing. This will likely continue as the academic standards for previously nontraditional information formats are established. As the availability and predominance of 3D data increases, so too will the pathways and procedures needed to access it. In general, record of who is accessing 3D data will not affect the long term veracity of said data but may be helpful to content creators and archivists, along the lines of existing circulation statistics, download counts, and page views. This may be collected in formats already in use within integrated library systems and website analytics. For future usage of 3D data, it may be worth considering whether different modes of access should be recorded and if that record may differ from that gathered for more commonly sought digital objects such as journal articles and images held in online databases.

Recommendations (3D-specific and/or general)

In general, the recommendations listed below were conceived to help stakeholders at different levels of 3D creation and practice. As the potential consumers of 3D data come from a variety of areas of expertise and backgrounds, care should be taken throughout to utilize common and standardized terminology wherever possible, ensuring the usage of both general keywords along with more specialized and technical creation data, in order to make the end product more discoverable by both 3D specialists and the general public.

Good practices, in this case, are what will be the most useful for the casual user who will need to find 3D data or a 3D model that suits their needs. At bare minimum, there should be sufficient citation and keyword metadata for an individual not related to the original creation to be able to find the data using common search parameters in search engines or library discovery layers as discussed in previous sections of this chapter. This may include but is not limited to: original item data, method of creation, creation or publication date, author or creator information, general subject terms including such information as area of study, general time period in history, reason for creation,[xli] possible derivative chains, and where the data can be accessed. For example, Digital Morphology [DigiMorph] recommends the following citation format:

```
Author/creator, [creation] year, “title of dataset,” name of the holding
collection. Access date and accessed URL.
```

As many casual users are currently seeking to consume 3D data through specific modes, file intent or purpose is also important at this level. This currently includes printability on a 3D printer [potentially including a necessary derivative or accommodation to afford this], usability within virtual reality, and access through a virtual viewer, either web-based or software driven.

[xli] Some examples may include archival rendering, teaching or study aids, historical recreation based on available sources, etc.

Better practices apply to most researchers, many of whom are not 3D practitioners themselves but are often familiar with common academic processes and information gathering behaviors. To best assist these users, whenever possible, globally unique identifiers [GUIDs] such as Digital Object Identifiers [DOIs] or Archival Resource Keys [ARKs] should be assigned to ensure a level of specificity in discovery on par with accession numbers like those utilized in library catalogs and large databases like WorldCat, a worldwide library database managed by the Online Computer Library Center [OCLC]. These GUIDs allow consumers of all types the ability to find a specific item with a minimal amount of information. Many of these users may also be looking for 3D data or models based on basic creation data or mode of creation--potentially photogrammetry, LIDAR capture, or CAD--but may not have the skills to create this data themselves or to parse more advanced capture data. Basic identifiers for available derivatives may also be useful to distinguish between versions, such as the articulated woolly mammoth model designer Teraoka Gensyou created from the Smithsonian Digitization model of a woolly mammoth skeleton.[16]

Best practices for access and reuse of data may be the most useful to 3D specialists and information and/or resource sharing specialists assisting researchers in locating 3D data as it involves the usage of highly specific data not necessarily relevant to lower intensity users. As it is not currently possible to stream high quality 3D data over the web due to the significant file sizes, web-based viewers of 3D data compress that data out of necessity. Resolution and compression data, as well as adaptive and scalable considerations, may be key in how certain 3D specialists interact with and verify that a given dataset will meet their needs and may give integral insight into how that data may be reused. Users at this level may find it useful to search by detailed collection data relevant to the raw data captured, including but not limited to information about photogrammetric sets, LIDAR settings, or poly count. Therefore, it is worth considering whether there should be multiple levels of GUIDs or whether there should be a single resolver for all iterations of a given set of 3D data. While a DOI for 2D data is used for accessibility, best practices for DOIs and access data for 3D data may need to go farther, potentially indicating compression rates, creation history, provenance, and more.

Good

Prioritizes discoverability by non-practitioners, enough data to cite the model/dataset:

- *Author(s)/creator(s)*
- *Creation/publication year*
- *Title of dataset/model*
- *Name of the holding collection [if applicable]*
- *Original item data [if generated from a scan of an object]*
- *Source material [if **born digital**[xlii]/recreation]*
- *File type*
- *File intent or purpose [printability vs virtual reality vs virtual view]*

Better

All of the above data plus:

- *Globally unique identifier [GUID] such as DOI or ARK for master data set*
- *Basic Creation data, mode of creation [photogrammetry, LIDAR, manual modeling, etc]*
- *Basic identifiers for available derivatives*

Best

[xlii] See Glossary

All of the above data plus:
- *Detailed collection data relevant to raw data [see Creation]*
- *Synthesis method to create master model*
- *Delivery methods for available derivatives, including relevant resolution and compression data*
- *Evaluate whether specific derivatives and/or delivery methods need a separate GUID.*

Archive

Metadata in the Archive stage is associated with the maintenance and preservation of 3D data. It is important to document the file formats, creation environments, and required software and hardware to ensure long-term access and reusability. Archival preservation of 3D data will often require consultation with a **digital preservation**[xliii] specialist at the beginning stages of the data creation process.

Stakeholders

Archivists and repositories should take stock of these requirements and recommendations as they manage much of the transition between distribution and preservation. Creators donating or submitting their work should also consider the preservation metadata of their data. Finally, those looking to reuse datasets that are older, or in unfamiliar formats will require much of the information recommended in this section to be able to gauge how, or whether or not they will be able to access preserved data.

Survey

To address gaps in preservation metadata practices, this paper uses the PREMIS Data Dictionary as a guide.[17] Much of the metadata specified in PREMIS shows how heavily dependent preservation metadata is on sufficient creation metadata. Based on our first survey, we found that practices, uses, and workflows for 3D data varied greatly, and we had no real responses indicating developed preservation metadata (as described later in this section's recommendations section). This complicates the development of digital preservation workflows, as each type of 3D data needed differing levels of information to support reuse. Based on this information, we focused on better understanding existing file formats and technical environment requirements (hardware and software needed to use 3D data sets). Regarding existing formats, respondents indicated a number of file formats used in various projects of varying size, scope, and purpose. There were twenty different formats listed within our limited survey responses alone, with each of these formats experiencing varying levels of support, maintenance, and platform compatibility. Not all formats were 3D specific. .pdf, .ppt, .txt, and .tif were among the non-3D file formats listed. Regarding environment, variation once again characterized the responses. Memory requirements and modes of storage vary with method of creation, project size, project type, and the creator's access to resources. Memory ranging from megabytes to multiple gigabytes and cloud storage vs. local storage were present in our limited samples across the two surveys.

Implications

The presence of multiple file types makes sense when considering the state of digital preservation. The work of preserving digital materials is still developing and ever-changing. Aside from

---

[xliii] See Glossary

basic storage and maintenance questions like those around server space and data degradation, there are additional concerns regarding the rapid change of preferred, utilized formats. Formats from larger, established companies, like Adobe and Microsoft, have the wide use required to ensure continued demand for access. They also have the support and funding required to meet that demand by supporting continued access to their older formats via conversion options. Their wide use also makes it profitable, or simply useful, for independent developers to invest in the development of tools that support the conversion and viewing of obsolete formats. There is also more precedent in preservation efforts for these older, established products (Word, Illustrator, Powerpoint, etc.). With that in mind, metadata becomes increasingly important because the data is removed from its original context. This means that the metadata will need to compensate for this loss of context, particularly in the case of datasets where reuse requires the ability to replicate creation. Particular attention will have to be paid to retaining detailed creation metadata.

Aside from file type and technical/environment requirements, considerations around data derivatives for publication and sharing were discussed. These derivatives should be treated as unique datasets with their own creation metadata and additional fields linking them to, and describing their relationship with, the dataset upon which they are based.

Recommendations (3D-specific and/or general)

Before any formal metadata is created, it is recommended that creators work with an archivist to make early decisions regarding the preservation needs of their work. As previously stated, this chapter recommends the utilization of PREMIS. The exact PREMIS entities and semantic units used will depend on the type of 3D data as well as its intended purpose, as PREMIS does allow for a high degree of flexibility that would support the more granular metadata models utilized by institutions like the Smithsonian.[18] Special attention should be paid to the “Special Topics” portion of PREMIS. Metadata around environment and object characteristics are particularly important to 3D, especially if data has been compressed or is platform/operating system specific. The recommended metadata from the “Create” and “Manage” portions of this chapter should be carried over into this one. Recommendations in this section will be based upon the infrastructure and support available to stakeholders. Digital preservation typically requires a digital preservation specialist, server space, and a system that supports the metadata recommended in this chapter. As a result, recommendations under “Good” will be for those utilizing a proprietary content management system with limited metadata flexibility. Recommendations under “Better” will based on access to content management systems that support 3D, but do not specialize in it. Recommendations under “Best” will be based on access to a content management system designed to host 3D data.

**items with an asterisk are direct semantic units in PREMIS; some of these have multiple parts, that is apparent when reading the PREMIS entity description*

Good
In order to capture and retain metadata in content management systems that do not support the more granular metadata required for reuse, it is recommended that unsupported essential metadata be kept in an accompanying easily supported file format such as .txt or .csv. It is also recommended that metadata fields be encoded into a commonly used markup language (like XML) when possible. This will make it easier to crosswalk and edit metadata should ingestion into a more supportive repository occur. At bare minimum, in addition to the original “Create” and “Manage” metadata, archivists should include:

-Repository ID
-*objectIdentifier (ISBN, DOI, etc.)
-*objectCategory
-*significantProperties (not required, but strongly recommended that the reader keep this unit in mind)
-*objectCharacteristics
-*formatDesignation
-*creatingApplicationName
-*creatingApplicationVersion
-*contentLocationValue
-Related objects
-Event information (metadata on compression, edits, or other actions taken)
-Rights information
-Accessibility information
  -Alt Text
  -Long description
  -Accessible format/information location (alternative accessible format or accessibility policy information)

Better/Best

The capabilities of the content management system being used will heavily determine the feasibility of incorporating additional metadata entities. Systems that can automatically capture metadata regarding event information and file format will be necessary to meet the more extensive requirements in PREMIS for 3D.

-More extensive agent metadata (see agent entity section of PREMIS)
-More extensive environment, as well as other "Special Topics" metadata (see "Special Topics" section of PREMIS
-Separate event records for tracking preservation activities

Gap Analysis/Future Work

As 3D creation, storage, and usage practices are still maturing, much of what is discussed in this chapter will continue to evolve as the practices solidify. Both software and hardware involved in the creation of 3D data is changing rapidly which makes uniform suggestions difficult. The situation is further complicated because there is evolution on two fronts, the technology itself and its use. Both scholastic and industry approaches to the creation and use of 3D data is changing heavily. Initiatives like the Library of Congress' Born to Be 3D effort, LIB3DVR, and CS3DP are working to both contextualize and focus academic approaches for 3D data while organizations like MorphoSource, Cultural Heritage Imaging, IIIF, and the Smithsonian are working to clarify both 3D creation standards and/or ease of access. While industry is not represented in the survey results for this chapter, it is important to note that there was industry participation at the second CS3DP forum and industry efforts will impact how 3D data management practices progress.

3D models offer a powerful and information rich way to document the physical condition and appearance of objects and environment as they existed in a moment in time. Aside from being a visual and metrological record, 3D models provide a scaffold for deeper documentation and compelling storytelling about the subjects they describe. 3D models can be annotated in a variety of ways, such as connecting text and other media to discrete points or regions of interest (either on the models surface or volumetrically within the models world space). When the primary goal of a 3D model is to document its subject, this information can be as important as the model itself. Citations and annotations of said data are still evolving and will continue to do so as scholars find more ways to take advantage of the opportunities this format presents. As the technology develops, so does the creativity in the community's use of it.

As 3D expands in its use, there are additional legal and ethical concerns that need to be addressed as well. One in particular is accessibility for disabled users, especially when being used in educational settings. While there are a myriad of issues that need to be addressed in order for 3D to be accessible, this chapter has shown the bare minimum metadata that should accompany 3D models when possible (alt text, long description, alternative format/accessibility information). However, this minimal metadata does not support equitable use and will need to be expanded upon and diversified.[xliv] As is the case with other categories of metadata, accessibility metadata needs will vary in different disciplines and industries and will need to convey both purpose and content. For example, functionality with screen readers will require more automated and nuanced orientation information. Having a version of 3D data that can be 3D printed, viewed, and manipulated in a browser, examined in layers so that textures and other features can be described, and/or examined and communicated as raw data are all potential approaches that will also require more effective metadata. This also means that there will need to be effective validation tools and audit workflows in place. Each field and discipline will have to work to determine what equitable use looks like for its disabled members and begin the process of determining what technology and metadata is needed. In order to truly support reuse for disabled users, avoid costly remediation, and foster the necessary cultural change required for buy-in, these issues should be addressed sooner rather than later.

Conclusion: Summary Recommendations

The following table summarizes the metadata practices recommended at PIPs through each stage of the digital asset lifecycle. There is no single metadata standard that encompasses all recommended fields or fills all the specific needs for 3D model metadata but metadata standards such as Dublin Core, PREMIS, and VRA Core can address portions of the recommended metadata fields. In terms of academic research data, metadata recommendations should adhere to the basic principles of FAIRness[19]. Additionally, domain specific work is in progress within groups such as Audubon Core, which may further provide 3D-specific field options in the future. Further community driven work is needed, however, to develop and refine a metadata standard that is truly cross disciplinary, cross

[xliv] Equitable use is a principle of Universal Design. In order to be considered as having provided equitable use, VR must: 1) Provide the same means of use for all users: identical whenever possible; equivalent when not; 2) Avoid segregating or stigmatizing any users; and 3) Provisions for privacy, security, and safety should be equally available to all users.; United States Access Board, "Principles of Universal Design," accessed on 7 October 2020, https://www.access-board.gov/guidelines-and-standards/communications-and-it/26-255-guidelines/825-principles-of-universal-design

modal and widely interoperable. These recommendations identify necessary categories of metadata that are required in the emerging adaptable, flexible standard.

Aggregated Table of GBB Recommendations

	Good	Better	Best
Create	Document and use a folder structure and file naming convention to organize source data and 3D models for management purposes. For example, group data and models by digitization project, collection, etc. Include the following information in README.txt files, CSV files, spread sheet or similar. Project level information: - Project name - Project identifier - Project date - Description/ abstract - Project authors/creators - Stakeholders/ contributors - Project rights information Source data information, if a reality capture model or sources-based model: - Method of creation - Creation date - Information identifying real world object or environment - Creators 3D model information: - Subjects - Creator - Geometry information - Textures (what	Include the following additional information in README.txt files, CSV files, spreadsheet or similar. Source information for reality capture or source based models - Resolvable GUID for records for subject or sources - Data source (capture device make and model) - Georeference information if applicable Source data creation information - Capture device make and model - Capture event details 3D model processing information - Source data used - Software used Documentation of capture and processing workflows	Include the following additional information in README.txt files, CSV files, spreadsheet or similar. 3D model processing information - Detailed steps and log outputs Use standardized metadata properties to define project, file, and 3D model properties listed above

	types of UV maps are available) - Materials (If any material properties are applied to a model)		

Manage	Group 3D models together as project or collection for access and management purposes.
	
	Include following project level data:
	- Project name
	- Project identifier
	- Project date
	- Description/ abstract
	- File/data types
	- Subjects
	- Project authors/creators
	- Stakeholders/ contributors
	- Project copyright information
	Supply copyright information at highest level possible (project/collection/object/file)
	Using available tools,[xlv] document the following properties from files in a structured file type such as a README.txt file:
	- File-specific:
	  - Filename
	  - Checksum
	  - File format
	  - File size
	  - File date
		Use digital repository/digital asset management system with access capabilities to define permissions and make project/collection discoverable.
		
		Store project, file, and 3D model properties from Good column with 3D models in digital repository / digital asset management system.
			Use rightsstatements.org statements and CreativeCommons licenses to define standardized access and reuse levels.
			
			Use standardized metadata properties to define project, file, and 3D model properties listed in Good and Better columns.

[xlv] Example of available tools include DROID (https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/) and FFProbe (https://ffmpeg.org/ffprobe.html) but there are others available to also do this work.

	creation - File date modified - File version - 3D model-specific: - Scale - Number of vertices - Geometry type[xlvi] - Number of faces - Number of points - CT data-specific: - Number of		

[xlvi] For example, polygonal quads, polygonal tris, polygonal ngons.

[1] London Charter, "Glossary," accessed on 28 August 2020. http://www.londoncharter.org/glossary.html

[2] Filecamp, "Digital Asset Management: What is it and why do you need it?" accessed on 28 August 2020. https://filecamp.com/blog/what-is-digital-asset-management/

[3] Digital Curation Centre, "The DCC Curation Lifecycle Model," accessed on 2 October 2020. https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf

[4] London Charter.

[5] N. A. Matthews, *Aerial and Close-range Photogrammetric Technology: Providing Resource Documentation, Interpretation, and Preservation*, Technical Note 428 (Denver, Colorado: U.S. Department of the Interior, Bureau of Land Management, National Operations Center, 2008).; Cultural Heritage Imaging, "Photogrammetry: Practical, Scientific Use of Photogrammetry," accessed on 28 August 2020, https://vimeo.com/channels/practicalphotogrammetry

[6] Cultural Heritage Imaging, "Digital Lab Notebook," accessed on 28 August 2020, http://culturalheritageimaging.org/Technologies/Digital_Lab_Notebook/

[7] Texas Tech University Libraries, SHAPES, accessed on 28 August 2020, https://shapes.lib.ttu.edu/

[8] Library of Congress, "Standards," accessed 28 August 2020, https://www.loc.gov/librarians/standards; Library of Congress, "Recommended Formats Statement," accessed 28 August 2020, https://www.loc.gov/preservation/resources/rfs/TOC.html

[9] Samvera, main page, accessed on 28 August 2020, https://samvera.org; Samvera, "Technical Metadata Application Profile," accessed on 28 August 2020, https://wiki.duraspace.org/display/samvera/Technical+Metadata+Application+Profile

[10] Smithsonian Digitization Program Office, "Smithsonian 3D Metadata Model," accessed 28 August 2020, https://dpo.si.edu/blog/smithsonian-3d-metadata-model; Archaeology Data Service/Digital Antiquity, "Guides to Good Practice," accessed 28 August 2020, http://guides.archaeologydataservice.ac.uk/g2gp/3d_3-3

[11] Sketchfab, main page, accessed 7 October 2020, https://sketchfab.com/

[12] NIH 3D Print Exchange, main page, accessed 7 October 2020, https://3dprint.nih.gov/

[14] FAIR Data Principles https://www.force11.org/group/fairgroup/fairprinciples

[15] Don R. Swanson, "Searching Natural Language Text by Computer." Science, New Series, 132, no. 3434 (1960): 1099-104, https://www.jstor.org/stable/1706747.

[16] Smithsonian 3D Digitization, "Articulated Wooly Mammoth," accessed 7 February 2019, https://3d.si.edu/explorer/articulated-woolly-mammoth; Teraoka Gensyou, "Articulated Woolly Mammoth Manga," *SI Digi Blog*, accessed 7 October 2020, https://dpo.si.edu/blog/articulated-woolly-mammoth-manga

[17] Library of Congress, "PREMIS Data Dictionary for Preservation Metadata, Version 3.0," accessed 7 October 2020, http://www.loc.gov/standards/premis/v3/

[18] DPO, "Smithsonian 3D Metadata Model," *SI Digi Blog*, accessed 7 October 2020, https://dpo.si.edu/blog/smithsonian-3d-metadata-model

[19] FAIR Data Principles, https://www.force11.org/group/fairgroup/fairprinciples

	- slices - Spacing between slices		

Distribution and Publishing	For casual use, users need to understand the format and intended use of the model. Intended use will usually dictate whether a model is suitable for printing, virtual reality, or other purposes. For an example, see Sketchfab.
	
	- System Unique Identifier
	- Author(s)/ Creator(s)
	- File type
	- File size
	- Geometry information (quads, triangles, etc.)
	- Vertices
	- Textures/Materials
	- Description
	Accessibility information
	- Alt Text
	- Long description
	- Accessible format/information location (alternative accessible format or accessibility policy information)
		3D data meant to act as an illustration in scholarship should include the data listed in the Good column in addition to the following recommended fields. For an example see the NIH 3D Print Exchange.
		
		- Link to original source material OR formal name and information of original item (scientific name, formal name, etc.)
		- All processing done (textures added, artistic liberties taken)
		- DOI/PID
		- Associated works
		- Citation information
			Datasets designed to be used as research data for fields and industries that are dependent upon accurate standards for reuse should include the data listed in the Good and Better columns in addition to the following fields. For an example, see Morphosource.
			
			- Detailed Capture Metadata (amperage, axis resolution, scanner resolutions, exposure time, wattage, etc.)
			- Project information:
			  - Funding information
			  - Affiliated organizations
			- Technician information
			- Checksum

Access and Reuse	Prioritize discoverability by non-practitioners, enough data to cite the model/dataset: - Author(s)/ creator(s) - Creation/	All of the data listed in the Good column plus: - Globally unique identifier [GUID] such as DOI or ARK for master data set - Basic Creation data,	All of the data listed in the Good and Better columns plus: - Detailed collection data relevant to raw data [see Creation] - Synthesis method to create master model

	publication year - Title of dataset/model - Name of the holding collection [if applicable] - Original item data [if generated from a scan of an object] - Source material [if born digital/recreation] - File type - File intent or purpose [printability vs virtual reality vs virtual view]	mode of creation [photogrammetry, LIDAR, manual modeling, etc] - Basic identifiers for available derivatives	- Delivery methods for available derivatives, including relevant resolution and compression data - Evaluate whether specific derivatives and/or delivery methods need a separate GUID.
Archive	In order to capture and retain metadata in content management systems that do not support the more granular metadata required for reuse, it is recommended that unsupported essential metadata be kept in an accompanying easily supported file format such as .txt or .csv. It is also recommended that metadata fields be encoded into a commonly used markup language (like XML) when possible. This will make it easier to crosswalk and edit metadata should ingestion into a more supportive repository occur. At bare minimum, in addition to the original "Create" and "Manage" metadata, archivists should include:	The capabilities of the content management system being used will heavily determine the feasibility of incorporating additional metadata entities. Systems that can automatically capture metadata regarding event information and file format will be necessary to meet the more extensive requirements in PREMIS for 3D. - More extensive agent metadata (see agent entity section of PREMIS) - More extensive environment, as well as other "Special Topics" metadata (see "Special Topics" section of PREMIS - Separate event records for tracking	See Better column

	- Repository ID
	- *objectIdentifier (ISBN, DOI, etc.)
	- *objectCategory
	- *significantProperties (not required, but strongly recommended that the reader keep this unit in mind)
	- *objectCharacteristics
	- *formatDesignation
	- *creatingApplicationName
	- *creatingApplicationVersion
	- *contentLocationValue
	- Related objects
	- Event information (metadata on compression, edits, or other actions taken)
	- Rights information
	- Accessibility information
	    - Alt Text
	    - Long description
	    - Accessible format/information location (alternative accessible format or accessibility policy information)
		preservation activities	

**items with an asterisk are direct semantic units in PREMIS; some of these have multiple parts, that is apparent when reading the PREMIS entity description*

Acknowledgements

The CS3DP Metadata Working Group would like to thank all survey participants and acknowledge the contributions of Andrea Thomer in helping to conduct Survey 1 and gather responses from participants.

Bibliography

Archaeology Data Service/Digital Antiquity. “Guides to Good Practice.” Accessed 28 August 2020, http://guides.archaeologydataservice.ac.uk/g2gp/3d_3-3

Cultural Heritage Imaging. “Digital Lab Notebook.” Accessed on 28 August 2020, http://culturalheritageimaging.org/Technologies/Digital_Lab_Notebook/

Cultural Heritage Imaging. “Photogrammetry: Practical, Scientific Use of Photogrammetry.” Accessed on 28 August 2020, https://vimeo.com/channels/practicalphotogrammetry

Digital Curation Centre. “The DCC Curation Lifecycle Model.” Accessed on 2 October 2020, https://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf

Digital Public Library of America. Main page. Accessed 7 October 2020, https://dp.la/

DPO. “Smithsonian 3D Metadata Model.” *SI Digi Blog*. Accessed 7 October 2020, https://dpo.si.edu/blog/smithsonian-3d-metadata-model

Dublin Core Metadata Initiative. “Dublin Core.” Accessed on 7 October 2020, https://www.dublincore.org/specifications/dublin-core/

Europeana. Main page. Accessed 7 October 2020, https://www.europeana.eu/en

FFmpeg. “ffprobe Documentation.” Accessed 7 October 2020, https://ffmpeg.org/ffprobe.html

Filecamp. “Digital Asset Management: What is it and why do you need it?” Accessed on 28 August 2020, https://filecamp.com/blog/what-is-digital-asset-management/

Gensyou, Teraoka. “Articulated Woolly Mammoth Manga.” *SI Digi Blog*. Accessed 7 October 2020, https://dpo.si.edu/blog/articulated-woolly-mammoth-manga

Library of Congress. “MARC Standards.” Accessed on 7 October 2020, https://www.loc.gov/marc/

Library of Congress. “MODS: Uses and Features.” Accessed on 7 October 2020, http://www.loc.gov/standards/mods/mods-overview.html

Library of Congress. “PREMIS Data Dictionary for Preservation Metadata, Version 3.0.” Accessed 7 October 2020, http://www.loc.gov/standards/premis/v3/

Library of Congress. “Recommended Formats Statement.” Accessed 28 August 2020, https://www.loc.gov/preservation/resources/rfs/TOC.html

Library of Congress. “Standards.” Accessed 28 August 2020, https://www.loc.gov/librarians/standards

London Charter. “Glossary.” Accessed on 28 August 2020, http://www.londoncharter.org/glossary.html

Matthews, N. A.. *Aerial and Close-range Photogrammetric Technology: Providing Resource Documentation, Interpretation, and Preservation*. Technical Note 428 (Denver, Colorado: U.S. Department of the Interior, Bureau of Land Management, National Operations Center, 2008).

Morphosource. “Getting Started.” Accessed 7 October 2020, https://www.morphosource.org/

National Archives. “Download DROID: file format identification tool.” Accessed 7 October 2020, https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

NIH 3D Print Exchange. Main page. Accessed 7 October 2020, https://3dprint.nih.gov/

Resource Description & Access [RDA] Toolkit. “About.” Accessed on 7 October 2020, https://www.rdatoolkit.org/about
Samvera. Main page. Accessed on 28 August 2020, https://samvera.org
Samvera. “Technical Metadata Application Profile.” Accessed on 28 August 2020, https://wiki.duraspace.org/display/samvera/Technical+Metadata+Application+Profile
Sketchfab. Main page. Accessed 7 October 2020, https://sketchfab.com/
Smithsonian 3D Digitization. “Articulated Wooly Mammoth.” Accessed 7 February 2019, https://3d.si.edu/explorer/articulated-woolly-mammoth
Smithsonian Digitization Program Office. “Smithsonian 3D Metadata Model.” Accessed 28 August 2020, https://dpo.si.edu/blog/smithsonian-3d-metadata-model
Swanson, Don R. "Searching Natural Language Text by Computer." Science, New Series, 132, no. 3434 (1960): 1099-104. https://www.jstor.org/stable/1706747.
Texas Tech University Libraries. *SHAPES*. Accessed on 28 August 2020, https://shapes.lib.ttu.edu/
United States Access Board. “Principles of Universal Design.” Accessed on 7 October 2020, https://www.access-board.gov/guidelines-and-standards/communications-and-it/26-255-guidelines/825-principles-of-universal-design

Endnotes

[13] Morphosource, “Getting Started,” accessed 7 October 2020, https://www.morphosource.org/