

UC Santa Cruz

Library Staff Presentations and Research

Title

Best Practices for 3D Data **Preservation**

Permalink

<https://escholarship.org/uc/item/2tf3c0w4>

Authors

Golubiewski-Davis, Kristina

Maisano, Jessica

McIntosh, Marcia

et al.

Publication Date

2021-11-17

Peer reviewed

Best Practices for 3D Data Preservation

This version of the chapter is a pre-print.

Full citation for the upcoming publication is:

Moore, Jennifer, Adam Rountrey, and Hannah Scates Kettler. 2022. 3D Data Creation to Curation: Community Standards for 3D Data Preservation. Association of College & Research Libraries.

This chapter pre-print is published under [CC-BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/) licensing.

Best Practices for 3D Data Preservation

Authors: Kristina Golubiewski-Davis, Jessica Maisano, Marcia McIntosh, Jennifer Moore, Kieron Niven, Will Rourk, Rebecca Snyder

Abstract

The CS3DP Preservation Best Practices working group has focused on creating a framework that can be applied across multiple 3D-data-generating modalities and tiers of implementation, from national to local governmental, educational, and private entities. The primary focus in this chapter is on identifying **Preservation Intervention Points**ⁱ (PIPs) within project **workflows**ⁱⁱ, i.e., moments to stop and ask: “What files should I save?” and “What information should be recorded at this point?” The answers to these questions are project dependent and related to the purpose of generating the data, the target audience(s), **preservation**ⁱⁱⁱ needs, and creation modality. However, some general guidelines are provided for new and established practitioners to consider. This chapter breaks down the PIPs into four main areas: Planning; Collection and Creation; Processing; and Curation and Long-Term Access. At each stage, we recommend levels of implementation to accommodate different levels of infrastructure, funding, and audience requirements. The chapter concludes with a discussion of **Good/Better/Best**^{iv} recommendations and is supported by an extensive appendix comprising case studies illustrating PIPs and the Good/Better/Best recommendations for a range of 3D-data-generating modalities.

Introduction

As a practice, **digital preservation**^v has deep connections with disciplines such as the library and archival sciences. Libraries and **archives**^{vi}, with their centuries-long history of experience, have become so ingrained within the consciousness of the people they serve that their inherent value is implied in their very names. Say “library”, especially within a museum or academic context, and no one would question why it is important or what purpose it serves. The challenge for digital preservation, as a relatively new field, is to continually demonstrate the value it provides to the communities it serves. It can be difficult to make a case to an institution’s administration as to why additional funds, staffing and other resources must be given to ensure long-term viability, discoverability and **access**^{vii} to digital resources. Often it is assumed that, because paper documentation is “so easy” to maintain over time, digital data should not be any different. Paper preservation seems effortless to a non-conservator simply because of the decades of work and research that have provided a strong foundation for it to become common practice. Digital resources are essential to the mission of contemporary organizations, society, and knowledge creators. Over time - *and many lost files later* - the institutional value of digital preservation will likewise become an assumption. Best practices for digital preservation provide an important foundation towards supporting this goal.

ⁱ See Glossary

ⁱⁱ See Glossary

ⁱⁱⁱ See Glossary

^{iv} See Glossary

^v See Glossary

^{vi} See Glossary

^{vii} See Glossary

Existing Standards

Digital preservation typically is divided into three major foci - Management, Technology, and Content - often with an accompanying diagram of a three-legged stool or a Venn diagram “triad”¹ to illustrate how all preservation activities are interlinked and interdependent. Indeed, the intention to provide access to and preserve digital materials long into the future is impossible without: institutional will, funding, clear policies and review cycles (Management); a well-designed and robust repository that is tailored to the targeted types of data and designated communities (Technology); and well-defined selection criteria, **metadata**^{viii} standards, and a deep understanding of formats (Content). Because digital preservation touches nearly every aspect of an organization, the task of converting that intent into a tangible plan of action can be daunting.

A group of archivists responsible for astronomical data recognized the need for preservation content management standards and formed the Consultative Committee for Space Data Systems. Their reference model for the Open Archive Information System (OAIS) has become an International Organization for Standardization (ISO) standard² and the basis for thinking about the management and preservation of digital materials among digital preservation practitioners. The companion, Audit and Certification of Trustworthy Digital Repositories³ (often referred to as TRAC) was designed to help institutions implement the particularly dense OAIS reference model and assess compliance through a checklist. The standard divides the OAIS model into three sections: Organizational Infrastructure; Digital Object Management; and Technical Infrastructure and Security Risk Management. The checklists require an institution to consider multiple factors when designing a repository or considering adding new functionality and/or support for new data types. More than simply focusing on software and storage, both OAIS and TRAC place great emphasis on clearly established policies, documentation of user needs, and long-term maintenance requirements. It is a misconception that digital preservation is primarily about technology and formats. While these are clearly important, sustainable preservation cannot occur without a robust organizational investment and long-term commitment. See Chapter 3, *Management and Storage of 3D Data*, for additional discussion.

PREservation Metadata: Implementation Strategies (PREMIS) is a working group focused on creating common metadata standards for preservation activities. The PREMIS Data Dictionary for Preservation Metadata⁴, currently in version 3.0, is designed to assist digital repositories in capturing vital information relating to digital preservation concerns. The best practices detailed in this chapter are compatible with the concepts detailed in the **data model**^{ix} and provide a means for implementing PREMIS for 3D data. See Chapter 3, *Management and Storage of 3D Data*, for a more in-depth discussion. Ideally, a **digital repository**^x would combine the OAIS model with PREMIS-compatible metadata standards.

Since 2016 the concept of FAIR data (data that is Findable, Accessible, Interoperable, and Reusable) has had a growing presence within digital preservation with the aim of highlighting the value and reuse potential of research data⁵. As with the three foci highlighted above, the FAIR principles are built around the three similar core areas (entities) of data, metadata, and management infrastructure⁶. While the FAIR principles are not a rigid, precise standard, they provide a backdrop against which projects, workflows, tools and systems can be assessed in order to ensure that data and metadata are accurately and consistently created, and in ways that are discoverable, accessible,

viii See Glossary

ix See Glossary

x See Glossary

and persistent. The FAIR principles relate directly to the Preservation Intervention Points (PIPs) framework discussed below, the latter framing these considerations within a data-creation-to-deposition workflow. To some extent, implementation of many of the FAIR principles is the responsibility of the archive or repository to which the data may ultimately be submitted. However, an awareness of these key considerations by data creators early within a project can only help to ensure that datasets - and the institutions in which they are deposited - are able to comply with these principles.

The goal of digital preservation is challenging enough for standard archival and library materials, which are well studied and understood as content types and formats. It becomes much more nebulous when trying to manage and preserve emerging digital forms, covering numerous approaches and producing a variety of potentially unstandardized open and proprietary data formats. This is particularly true for 3D data, from their raw stage to their fully processed/derived final form. As the name implies, the CS3DP group is dedicated to creating community standards for the preservation of 3D data. This chapter in particular considers 3D as a new content type that must be brought into existing OAIS best practices. These criteria and considerations are well established for other content types; the challenge is to modify and customize them for the unique properties and workflows associated with the major 3D data collection and processing methodologies, while remaining flexible enough to accommodate new 3D data collection modalities. These include assessment concerns by collection method/data types, retention schedule guidelines, evaluation criteria, discipline-specific standards, and the varied needs of designated communities. There can be no one solution that accommodates all institutions, all budgets, all data types, and all community needs. Questions like: “How often should I review file format viability?” “Should I maintain all the raw data long-term?” “How would my designated community expect to discover and receive content?” “What derived data and versions should be preserved?” will be answered by each institution according to their capabilities and requirements.

This chapter seeks to help institutions and individuals ask the vital questions and provides tiered solutions based on need and resources, using a PIPs approach that details key assessment points throughout the lifecycle of a 3D project, and poses key questions to consider. An institution's priorities, capabilities, selection criteria, designated community requirements, and available funding will all inform how these questions are answered and what recommended actions are taken. Baseline recommendations for preservation are offered in a Good/Better/Best format at the end of this chapter (tables 4.1, 4.2, and 4.3), with the intent that the PIPs framework will guide the implementation of preservation work. It is recommended that, at a minimum, data be saved in open-standard, non-proprietary, and human-readable file formats (e.g., ASCII text or XML-based formats), as this removes a dependency on specific software packages and allows greater flexibility for future access and migration. Throughout this chapter, the project lead is referred to as the decision maker. This choice was intentional to reflect the decisions that project leads often make early in the project, although this decision-making process could also be undertaken by a team or an institution.

Preservation Intervention Points

The basis for these ‘best practices’ recommendations is the framework of Preservation Intervention Points (PIPs). This framework comes from the Archaeology Data Service Venus Preservation Handbook⁷, expanded upon in the Archaeology Data Service Guides to Good Practice⁸. The basic premise of the PIPs framework is to identify and assess critical decision points in the data creation process and record those points within the context of their long-term implications for data preservation and reuse. Given the variety of 3D data acquisition methods discussed in this chapter

(and in this volume as a whole), the following sections identify the stages where PIPs are most likely to occur in any 3D data project, with specific examples of PIPs in the 3D data acquisition workflow of several Case Studies (Appendix X).

The following stages are considered most relevant for 3D data collection/creation workflows:

- Planning
- Collection/Creation
- Processing (Post-Acquisition Data Manipulation)
- Curation and Long-Term Access.

Ideally, when considering preservation of specific 3D data, the data collection/creation workflow will be mapped prior to beginning the process. Greater documentation of a specific project will help to identify which aspects of the workflow are worth preserving. 3D data collection/creation includes a wide variety of technologies and processes. The framework outlined in this chapter is intended to be flexible, allowing the project lead to implement the stages that are relevant to their project. Not all stages will be necessary for all projects. For example, a **born digital**^{xi} project may not include 3D data collection, but may include a phase of collecting other types of **source material**^{xii}. The sections below describe the generalities of the PIPs framework and some specific examples.

To see the PIPs framework in the context of an entire project, Case Studies are provided in Appendix X that document specific workflows for projects that use born digital, **photogrammetry**^{xiii}, X-ray CT, large-scale **laser scanning**^{xiv}, and **structured light**^{xv} scanning acquisition techniques. Each case study identifies the PIPs stages within that workflow and recommended actions for preservation. These examples also serve to illustrate how the PIPs framework can be adapted to fit specific institutional priorities and needs, as each project deals with these differently.

[insert PIP_visualization.png goes here. Caption: “Basic Preservation Intervention Points (PIPs) within a digital project. Image created in diagrams.net”]

Each of the 3D project stages below indicates a point at which the project lead identifies and records information for long-term preservation. For each preservation stage, we have identified a series of questions for project leads to consider and the preservation implications of the answers to those questions. Within the context of a project, PIPs occur when the project has reached a predetermined milestone or when the data/data format have been altered in an irreversible way. These are the points at which the process and any appropriate metadata should be recorded, together with an assessment of data format in light of long-term preservation and access (i.e., avoiding pathways that ultimately lock data into proprietary or inaccessible formats). The goal of the PIPs framework is to create a plan that accounts for preservation while being flexible enough to allow for pivoting within the project based on unforeseen variables. An outline of known PIPs considerations and documentation methods created at the beginning of the project will benefit the process and should be available to all members of the project group.

As project leads consider the PIPs and questions detailed below, they will need to make context-specific decisions that meet the needs of their project, their institution and their designated

^{xi} See Glossary

^{xii} See Glossary

^{xiii} See Glossary

^{xiv} See Glossary

^{xv} See Glossary

audiences. As a guideline, at the end of this chapter there is a set of Good/Better/Best recommendations for implementation; see Tables 4.1, 4.2, and 4.3.

Planning

The planning stage is the most important stage at which to consider preservation because this is when the scope of the project and requirements for long-term preservation are decided. The purpose of the project, expected reuse, and needs for dissemination and preservation will affect how information is created and preserved during the workflow. It is strongly recommended that the 3D data creator/collector consider all project stages early in the process, including determination of an appropriate long-term storage solution for the project data and consultation with repository representatives, if relevant. While some of the information can be recreated at a later point in time, some information can only be captured during the data creation stage (see Chapter 4, *Metadata Requirements for 3D Data: Create* section). As the preservation and reuse of the data can hinge on specific data format choices and whether or not metadata and **paradata**^{xvi} (see Chapter 6, *Accessing 3D Data*) were recorded at the appropriate time, the 3D data collector/creator should carefully consider how the questions below affect a specific project and workflow (see Table 1). Educational institutions often offer support for project planning and 3D data preservation, usually via institutional repositories, libraries, humanities centers, or technical support in various departments.

The planning stage also includes time spent preparing for 3D data collection/creation. Preparing the location, creating collection procedures, gathering equipment, and testing the conditions to determine the best collection protocol are all done prior to acquiring the first 3D data collection point. The Large-Scale Laser Scanning case study in Appendix X provides an example that includes complex planning factors, multiple experts, and limited access to the site prior to destruction. Any decisions or factors that affect the collection process should be documented in a consistent manner. Careful attention should be paid to file organization techniques, and unique identifiers should be used for projects that include 3D data creation for multiple objects. Documentation of both the process and the reasons for collection or data optimization methods can inform the choices and priorities determined during those stages if access to the material is restricted by time or other unknown factors.

Purpose of the data collection/creation

The primary focus of 3D data collection is to provide enhanced dimensional information for the documentation of an object beyond that which can be acquired through 2D capture. This information may include both external and internal 3D properties of the object. For cultural heritage subjects, the geometry and texture are vital for the historic record of physical places and things. Data associated with specific buildings or sites may have specific points associated with real-world coordinates. The original purpose of the data collection/creation will impact the preparation of derivatives of the 3D data. For example, projects intended to provide accurate surface measurements will require datasets with a high density of points/surface models with a higher polygon count and recorded unit size for reference, whereas those intended for 3D printable models will comprise non-manifold closed surfaces and will not require as high a polygon count as detailed research data. Likewise, it is more important to consider accurate color representation for models used to increase public engagement through online interaction than for data that will be used to take measurements. These considerations, taken early on, can help guide a practitioner in considering which data types and

^{xvi} See Glossary

formats to preserve to meet the needs of the data/model use, and also the potential reuse of the data. For example, if the data were originally created for a 3D printing project, but the project was created through 3D scanning that yielded high-quality original scans that could be useful for measurement projects, then the preservation of detailed metadata, descriptions of the scanning process, and files from intermediate steps would increase the likelihood of reuse of the project data and the ability to check for scientific **accuracy**^{xvii} (see Chapter 6, *Accessing 3D Data*). The case studies in Appendix X include data whose purposes include educational outreach (Photogrammetry), cultural heritage preservation by use of 3D documentation (Large-Scale Laser Scanning and Multi-Modal), visualization of data for academic narratives (Born Digital), scientific analysis (High-Resolution^{xviii} X-ray CT), and creation of a 3D assemblage of objects for academic study (Structured Light Scanning). Each case study offers an explanation for how and why each institution chose to record information based on the needs of the project and the capabilities of the institution.

Detailed metadata and paradata should be collected for any project (see Chapter 4, *Metadata Requirements for 3D Data*). The decision to collect and preserve higher quality data than what is required for a specific project should be made based on the ability of the research team to collect those data, and the ability of the institution to properly preserve those data. Tables 4.1, 4.2, and 4.3 offer baseline recommendations for which files to preserve based on the purpose of the project. Ultimately, the 3D data collected and preserved for a project will be a balance between the best practice recommendations and individual project needs.

Intended audiences

When creating 3D data, the acts of documenting the process and preserving the data are recommended but rarely done unless the data will be critically evaluated (e.g., by scientific peers) and/or reused. At a minimum, the audience consists of the individual doing the work and/or the client for whom the work is done. For these reasons, careful documentation and proper preservation will benefit continued work involving the 3D data. In the Chapter 6, *Accessing 3D Data*, the authors define six categories of audiences: Scholars/Researchers; Educators; Students; Museums/Public Outreach/Non-Government Organizations (NGOs); Professionals; and General User/Personal Interest. Each of these groups will have different needs regarding the level of documentation and access to raw data or intermediary files. For example, scholars/researchers typically require a high level of documentation, in-depth metadata, and access to the raw files in order to verify the accuracy of the data/model and incorporate relevant data into their own projects. Public outreach groups will require access to the final project products in open file formats that are easy to reuse across a variety of platforms. Considering the needs of the intended audience during the planning stage will identify important PIPs stages for recording information and determine the level of preservation and access appropriate for the project.

Repository selection

Long-term preservation of 3D data is best achieved by depositing the project into a repository, preferably a trusted digital repository⁹. In some cases, this type of preservation may not be possible due to limited resources or restrictions on the data, in which case an alternative may include making the 3D data publicly accessible via an open-access repository or online database and finding medium-term solutions for storing the data (repository and storage examples are provided in Tables 4.1, 4.2, and 4.3). The purpose of the project and funding requirements will help to determine the

^{xvii} See Glossary

^{xviii} See Glossary

appropriateness of long-term preservation and the breadth and depth of the data preserved. Frequently, grant-funded projects have an obligation to publish and preserve data in a publicly accessible manner. During the planning stage, it is important to identify what preservation strategy will be applied to the data.

Long-term preservation often requires the identification of a repository in which to deposit the data. As of this publication, there is no universally-accepted file format, agreed-upon set of metadata, or standardized input format for 3D metadata. Identifying the repository for a project early on allows the project lead to understand the restrictions and implications of using different repositories. For example, some repositories allow files to be uploaded in a nested folder structure while others only allow a single folder for all of the files. Repositories might have a file type preference for the final ingest or specify a format preference and structure for metadata collection (see Chapter 3, *Management and Storage of 3D Data*, for more information). Understanding the requirements and limitations of the intended repository early in the process can expedite the collection of metadata and organization/formatting of files by matching the collection strategy to the needs of the preservation strategy. In addition to ensuring that data to be preserved meet the requirements of the selected repository, it is important to verify that the repository itself is capable of storing and preserving data in the long term. Should the chosen repository be unable or unwilling to preserve the volume of 3D data, metadata, and paradata necessary to fully encapsulate the project, either another repository should be considered, or a secondary preservation solution should be identified for the additional data. The OAIS model describes a standardized framework in which repositories can operate and, building from this model, a number of standards, assessment and certification processes have been created in order to allow institutions to highlight themselves as trustworthy data repositories (e.g., the CoreTrustSeal¹⁰). Repository certification is also discussed in the *Certification* section of Chapter 3, *Management and Storage of 3D Data*, 3D data creators can prepare to discuss depositing their data with a repository staff person by providing answers to several questions including¹¹:

- What is the origin/context of the data?
- What form and format are/will the data be in?
- What is/will be the expected lifespan of the data?
- How could the data be reused and repurposed?
- How large is the dataset, and what is its rate of growth?
- Who are the potential audiences for the data?
- Who owns the data?
- Does the dataset include any sensitive information?
- What publications or discoveries have resulted from the data?

Table 1 provides some examples and potential answers to fundamental questions that should be addressed at the initial stage of planning for 3D data preservation within the data workflow.

TABLE 1: PLANNING STAGE CONSIDERATIONS

Question / Answer	Implications
What is the purpose of the data or model creation?	The purpose of the project will determine the ways in which the data should be preserved and made accessible. See Chapter 6, <i>Accessing 3D Data</i> for an in-depth discussion.
Visualization	The final version of the project (model, animation, etc.) will be the most important set of files to preserve. Self-contained formats are more likely to maintain integrity. When multiple files are required to reproduce the visualization, file structure and relationships between files should be clearly organized.
Data Analysis	The analytical procedure is crucial to preserve. Original 3D data and information about how those data were edited/analyzed to obtain the final results will enable others to understand and verify those results.
3D Printing	Files intended for 3D printing should be relatively small and in .stl or .obj format. While raw data may be useful in the long term, it is important to provide them in a format that is easily migrated to current technology. Accompanying metadata should include print settings appropriate for the model, or settings/equipment used for specific printouts when the model is used in scholarly publications. Any editing done to make the model printable (e.g., hole filling, decimation, smoothing) should be clearly documented.
Historic or Cultural Record	3D data documentation inherently provides dimension and may provide texture or color information of the subject. For cultural heritage and historical subjects these data provide crucial documentation for the historic record.
Who are the designated audiences?	Different designated audiences will have different needs regarding documentation and metadata. This will also affect which stages and formats of the data should be preserved. See Chapter 4, <i>Metadata Requirements for 3D Data</i> for an in-depth discussion of metadata documentation and Chapter 6, <i>Accessing 3D Data</i> , for definitions and discussion of audience types. The following list provides examples of academic audience user groups for whom 3D data can help support research and pedagogy.
Scholars and Researchers	Documentation should include as much information as needed to replicate the study, including both technical and field-specific metadata. The 3D data should be made available minimally to peer reviewers. Persistent identifiers ^{xix} (e.g., DOI - Digital Object Identifier ¹² and ORCID - Open Researcher and Contributor Identifier ¹³) will increase data reuse and citation after publication.
Educators	The 3D data should be easily accessible, provided in a usable format, and documented to the level of intended use. Instructional supplements should also be included.
Students	The 3D data should be easily accessible and provided in a usable format.
Museums / Public Outreach / Non-Government Organizations (NGOs)	3D data ownership and reuse policies must be addressed, and proper credit given for data/model generation and funding source.
Professionals	Depending on the professional use, considerations for other groups may also apply here. When applicable, the level of documentation should follow the recommendations of any relevant government regulations for that field as well as any policies internal to the institution; these may include limitations on long-

^{xix} See Glossary

	term preservation and data sharing and use.
General User / Personal Interest	Guidelines should be provided for attribution of the 3D data and restrictions on its use (see Chapter 5, <i>Copyright and Legal Issues Surrounding 3D Data</i>).
What rights or obligations exist to share the data?	Obligations to share data may come from grants or institutional policies. The rights of the data collector/creator for sharing 3D data will depend on the purpose and nature of the project. See Chapter 5, <i>Copyright and Legal Issues Surrounding 3D Data</i> for an in-depth discussion of rights.
Grant-funded Obligations	Publicly-funded grants increasingly include obligations for openly sharing data where appropriate. These obligations should be planned for when determining the long-term preservation and access needs of the project.
Cultural Sensitivity Access	In some cases, the objects being transformed into 3D data may be culturally sensitive. This must be taken into account when determining the appropriate way to share their data.
Restrictions by Object Owners	Objects that are not owned by the project lead or their institution may have restrictions on the use of their 3D digital representations. See Chapter 5, <i>Copyright and Legal Issues Surrounding 3D Data</i> , for more information.
Georeferenced Data	References to the geolocation of object-based data may be culturally sensitive even if the data themselves are not. Sites that are thought to contain artifacts of significant number or cultural value or sites that are underwater may be susceptible to looting. Sites with endangered species may be subject to over-collection. These concerns should be taken into account before sharing geolocation of object-based data.
What is the appropriate repository for the data?	At the time of this writing, there is no standard for how 3D data are accepted into a repository, with most repositories having their own workflow preferences. Identifying a repository early in the project allows the project lead to incorporate repository requirements (e.g., data formats and metadata) into the workflow. See Chapter 3, <i>Management and Storage of 3D Data</i> , for an in-depth discussion.
Institutional Repository	Institutional repositories may not be familiar with 3D data. Be sure to consult the repository contact to understand possible limitations of the repository and discuss preservation strategies.
Non-institutional Repository	Each repository will have a different cost structure and ingest mechanism. Identify these early for incorporation into the project workflow and funding requests.
No repository	Understand the ramifications of not putting the data into a repository. Identify other options for disseminating the data if long-term preservation is not possible due to funding limitations. As with repositories, identify the file formats and associated information needed to use those resources early in the project.

Collection / Creation

The Collection/Creation stage refers to the point at which the raw data are generated. The duration of this stage and the number/volume of files generated depends on the method used to create the 3D data. It is important to document instrument settings, environmental settings, and acquisition protocols that affect this process. When possible, preserve any file outputs that the equipment software might generate (e.g., 'README' files that record equipment settings). Also when

possible, collect the highest resolution data with the most complete metadata (see Chapter 4, *Metadata Requirements for 3D Data*, for specific guidelines) and paradata available for the project - extra data is better than missing data. This is particularly true if the real-world object being captured by the 3D data will not be easily accessible for additional analysis. In addition to metadata, the collection/creation stage should include documentation of both the standard methods followed and the ability to note when and for what reason deviations occurred. If intentions behind particular data collection methods were developed during the planning stage, these can be used to help justify deviations from the standard method. Early discussion of the purpose of collecting the data, the level of detail in the metadata and paradata for the project, and the capabilities of the repository to preserve those data long-term are crucial to ensuring the project meets the minimum best practices recommended in this volume.

TABLE 2: COLLECTION/CREATION STAGE CONSIDERATIONS

Question	Implications
What are the raw data?	The raw data should be recorded and saved. Refer back to the intended audience and purpose of collecting/creating the 3D data to determine if they should also be preserved long term. Projects that are of an academic nature or result in the destruction of the original object, such as archaeological digs, should always preserve the raw data.
Does the equipment/software have options for metadata export (embedded or sidecar^{xx})?	Some software and imaging equipment have options for exporting a sidecar document (such as an associated 'README' text file) that describes the settings of the equipment and/or records a series of settings that the software used. In some cases, this information can also be embedded into the file itself. This information should be retained and kept with the 3D data.
What are the file format options?	During 3D data creation and collection stages it is normal that raw data is saved in its native file format. In cases where that format is proprietary, a non-proprietary format should be identified and preserved as well, particularly if continued access to software cannot be guaranteed. Where possible, both proprietary and non-proprietary formats should be preserved long term for research audiences. Some proprietary raw data files are not just a single file but a main project file and associated dependencies; this file structure must be retained for preservation and archiving. Refer to the section "Good/Better/Best Recommendations for Implementation" for preservation strategies of proprietary and/or non-proprietary file assets depending on archival institution resources.
What is the physical acquisition setup?	When using a 3D data acquisition method, the physical environment, setup, and equipment affect the quality of the resulting data. Document this information.
At what steps in the process should one stop and consider: <ul style="list-style-type: none"> ● Do I save this? ● Do I document this? If so, how? 	Each 3D data acquisition method will have a unique process. Determining the steps at which files are saved/processing is documented will create a series of PIPs for that particular process. This information can then be used to create a protocol that includes the appropriate documentation points and methods. All documentation should be preserved.

^{xx} See Glossary

Processing (Post-Acquisition Data Manipulation)

After the data acquisition phase, many 3D imaging modalities include a stage of data processing, which involves any modifications to the data between their acquisition and the final product. For example, raw data for photogrammetry comprises the original photographic images. These images then go through a processing stage where geometry and texture are created, cleaned, and exported as a final object. Structured light and laser scanners create one or more sets of data that are then aligned, optimized, and merged. Data intended for 3D printing may require editing (e.g., filling of holes, decimation, etc.). All of these modifications after the raw data acquisition should be carefully documented (see the Structured Light Scanning case study in Appendix X for an example). When possible, protocols and policies should be created that can be shared to assist the audience in understanding what manipulation was done to the raw data to create the final 3D data or model. In the case of bespoke data optimization, the changes that the data have undergone should be described and documented to the best of the ability of the individual making those changes. It is critical to record data processing or correction parameters (as exemplified in the High-Resolution X-ray CT case study in Appendix X) so that the final result may be replicated or understood for research and scholarly applications. As with the collection/creation stage, a standard method of processing should be documented along with any deviations from that method.

TABLE 3: PROCESSING STAGE CONSIDERATIONS

Question	Implications
At what steps in the process should one stop and consider: <ul style="list-style-type: none"> • Do I save this? • Do I document this? If so, how? 	All modifications to the raw data should be documented and that documentation preserved, but some PIPs project files need only be kept as working files and may not be included in the final project. Refer back to the intended audience and purpose of the 3D data creation to determine which working files should be preserved long term (See Chapter 3, <i>Management and Storage of 3D Data: File States and Submission Packages</i>).
In what format must the data be for the intended repository?	Understanding the structure and requirements of the repository and how these relate to the format and documentation of the project will expedite the translation of 3D data and documentation for inclusion in the repository. To determine what these requirements might be, contact the intended repository (See Chapter 3, <i>Management and Storage of 3D Data</i> , for guidance on choosing a repository).
Who needs access to the intermediate-stage data, and for how long?	Not all intermediate-stage data need be preserved, but there may be value in retaining these data for future work. Determining the audience and need for intermediate-stage data will help define an appropriate storage location and who should have access.
How likely are these data versions to be needed, accessed, or reused?	If intermediate data are likely to be needed for reuse of the final 3D data product, they should be preserved in the final data package.

Curation and Long-Term Access (SIP, AIP, and DIP)

SIP (**Submission Information Package**^{xxi}), AIP (**Archival Information Package**^{xxii}), and DIP (**Dissemination Information Package**^{xxiii}) are conceptual terms used by the OAIS reference model (ISO 14721)¹⁴ to refer to packages of information/data that are submitted (SIP), archived (AIP) and disseminated (DIP) by an archival body or repository. While the requirements and internal workings of repositories vary dramatically depending on location and scope, the basic concept that data may be deposited in one form, stored and preserved by the repository in another, and then disseminated to users in yet another range of formats highlights the fact that there is neither a one-size-fits-all nor a permanent solution to digital preservation and dissemination. The inner workings of digital repositories are beyond the scope of this chapter (see Chapter 6, *Best Practices for 3D Data Preservation*, and Chapter 3, *Management and Storage of 3D Data*). However, it is worth bearing in mind early on in any project involving 3D data destined to be preserved and accessed in the long term that the data created may be better preserved in a variety of formats, and in order for others to understand and use these data, sufficient documentation of their collection/creation and processing should also be stored alongside them. As discussed above, the FAIR principles should inform many aspects of this stage, ensuring that data are made available, and in suitable formats for reuse, alongside consistent and well-structured metadata.

The SIP – the package prepared for submission to the intended repository – is the point at which the creation of the project archive ends and the long-term preservation process begins. *Prepared* is the key term here, as almost all archives and repositories have specific guidelines and requirements for the submission of data and may refuse to accept a dataset if it fails to meet these requirements. Generally speaking, these requirements fall into two main areas: the type of data being submitted (covering both the content and the format); and the documentation and metadata that accompanies the 3D data. Some archives may only accept specific types of data, others only certain data formats and file types, typically with restrictions on both. A key to successfully preserving 3D data is to identify the repository and their requirements early on so that compliance can easily be built into the project workflow, ensuring data is ‘archive ready’ at the project’s completion. A worst-case scenario is only becoming aware of these requirements as the project is drawing to a close, resulting in a lack of time and money – or physical/technical ability – to undertake the tasks required to make the 3D data suitable for deposit (e.g., format migration, data documentation, data cleaning).

Even if data are not to be deposited into a repository (not recommended), a final preparatory stage formalizing a “project archive” and taking into account some of the considerations highlighted in this chapter is a worthwhile activity and will help ensure that a dataset is coherent and complete. For many projects this will include a stage of data selection based on previously highlighted PIPs but should also include general tasks such as the removal of duplicate data or multiple versions of files, ensuring filenames and directory structures are consistent, and making sure that there are no barriers to long-term preservation and access (e.g., technical, legal, or ethical restrictions or considerations).

A detailed early consideration of the software used in a project and the resulting file formats employed for data storage may bring to light future issues regarding long-term access and sustainability. In general, most approaches to digital preservation favor open-standard, non-proprietary, and human-readable file formats (e.g., ASCII text or XML-based formats), as this removes a dependency on specific software packages and allows greater flexibility for future access and migration. The sustainability of digital formats is discussed in detail on the Library of Congress Digital Preservation web pages¹⁵, and general Good Practice guides such as the Digital Curation Centre’s ‘Five steps to decide what data to keep’ provide a useful checklist¹⁶.

^{xxi} See Glossary

^{xxii} See Glossary

^{xxiii} See Glossary

Once the 3D data have been successfully packaged and ingested by the repository, an AIP can be generated. The AIP is the full package of data and metadata that forms the basis of the archival dataset. In OAIS technical terms, the AIP should consist “of the *Content Information* and the associated *Preservation Description Information* (PDI), which is preserved within an OAIS”¹⁷. In practice this consists of the original 3D data submitted for archiving, in a format suitable for long-term preservation, alongside the necessary information/metadata to aid preservation and to provide key details on **provenance**^{xxiv}, context, **fixity**^{xxv}, and so on. While the types of documentation that can be included in the PDI vary depending on the data and project type [see the Digital Preservation Coalition’s (DPC) Wiki page on ‘Preservation Description Information’¹⁸ for examples], detailed descriptions of provenance and context (e.g., recording methodology, instrument set-up, software processing, and intention), at various stages or PIPs, can be included here to allow a greater understanding of the 3D data, their purpose, and limitations. Data relationships, both between data within the package and to external elements, is key here, as the repository itself may need to repackage data for preservation purposes. A logical and structured directory system can make understanding a 3D dataset easier and make explicit the relationships between raw and processed elements. It is recommended, however, that such relationships are also documented elsewhere, and documentation is discussed in more detail in the following section.

Within a repository, additional metadata documenting any processing that the repository has undertaken on the 3D dataset will be created during ingest or migration. This might include format migrations, corrections to data, renaming, or restructuring, and aims to provide a chain of custody for the dataset (see Chapter 4, *Metadata Requirements for 3D Data: Distribute, Retrieve, and Archive* sections). Again, even for 3D data not formally deposited within a repository, the recording of edits made to those data beyond the scope of the original project is important to allow for derivatives and later versions of data to be traced back to their original forms.

In addition to the preservation data stored as the AIP, a DIP is created by the repository to allow access (at whatever level is appropriate) to the 3D data, usually as a download. In practice this may include all the submitted data, or it may be a discrete subset (e.g., processed data), or data derived from the AIP by the archive itself to provide different formats or different resolutions from those originally deposited. Such packages may also take into consideration other issues such as copyright restrictions or time-limited embargoes (see Chapter 3, *Management and Storage of 3D Data: Embargoes* section and see Chapter 5, *Copyright and Legal Issues Surrounding 3D Data*). While the DIP is largely specific to the repository, it is a worthwhile concept to consider when creating the SIP. If data contained within the SIP are simply different versions of the same data, then these relationships should be made clear to aid both the archive and any end user’s access and reuse of the data in the correct context. Likewise, any restrictions should be clearly documented so that data can be withheld from or disseminated to the correct users.

Documentation

A major goal of the PIPs framework is to create points of documentation for the 3D data collection/creation workflow. This documentation should be complete enough that someone can understand the work done to the data at any point in the process. This includes both individuals

^{xxiv} See Glossary

^{xxv} See Glossary

looking at the archived data as well as individuals who might join in the middle of a project. The documentation structure should be compatible with institutional, project, and preservation needs. Certain repositories or institutions will request documentation in particular file formats or following specific guidelines. Additionally, different 3D imaging modalities may require different types of accompanying documentation. Without proper documentation, 3D data are useful as visualizations, but lack the necessary information for reuse by scholars/researchers (see Chapter 6, *Accessing 3D Data: Reuse* section). The form that documentation takes varies based on project needs, but may include:

- Embedded or sidecar metadata exported from 3D acquisition equipment software
- Incorporation of metadata into file naming conventions
- Information pertaining to file relationships needed to properly open the data
- Associated spreadsheet data with documented information
- Published written protocols for replication of method
- Associated README text files that describe any of the above information

The method(s) of in-process and preservation-level documentation should be determined during the planning stage. The format of the data collection can be either structured or unstructured. Structured data, such as a spreadsheet, are appropriate for information that can be searched in a database, while unstructured data, such as a text file, are appropriate for recording protocols or other narrative information about the project. A common document found in US standards is a README.txt file, also known as a codebook. This is an associated (sidecar) file that describes the collection/creation of the 3D data, the meanings of any column headings in spreadsheet data, rights management information, preferred citation attribution, explanations of file structure, and may also include protocols for processing the data including the software used (including version). Examples of how each stage mentioned above can be expressed in a **README file**^{xxvi} can be found in the X-ray CT and Structured Light case studies in Appendix X. The project lead should identify the appropriate standards for their field and region of the world in regards to how to format and name the README document. An example of good documentation in practice for photogrammetry is the Cultural Heritage Imaging Digital Lab Notebook.¹⁹

Documentation should also take into consideration the appropriate file formats for the project including 3D data formats, metadata storage, and associated process documentation. Proprietary formats are tied to specific software suites, are the closest version to the raw data, and contain metadata specific to the capture process. Proprietary formatted data usually reside in the project file for the **registration**^{xxvii} of datasets and usually contain metadata from the instruments that collected the data. For reproducibility of the data, whether in the long term or short, the information bound up within a proprietary formatted project file is vital as it can contain instrument settings and readings from various sensors in the data acquisition instrument, such as altimeter, GPS and inclinometer sensors. Also important to the project file are the settings that were used to acquire 3D data such as the resolution of data at the time of capture, for example the number of points captured per scan session or the **voxel**^{xxviii} element dimensions in a CT scan. Many datasets comprise multiple datasets that have undergone some process of registration to combine them to create one comprehensive dataset. Proprietary formatted project files can contain the data regarding the **precision**^{xxix} with which the data were registered as well as the method of registration, thus ensuring fidelity and accuracy of the 3D data in the documentation of a subject. The ability to export project-related metadata cannot

^{xxvi} See Glossary

^{xxvii} See Glossary

^{xxviii} See Glossary

^{xxix} See Glossary

always be assured, but when it is possible to do so, they should be exported to a text-based format such as a .txt or XML document, especially if it is not permissible to archive the proprietary formatted project files. The Large-Scale Laser Scanning case study in Appendix X is an example of a 3D data acquisition type that produces proprietary formatted project files such as those described here.

Non-proprietary formats offer a significant benefit in that they can be read without specific software suites and are therefore more accessible and easier to maintain file fidelity in an archive (see Good/Better/Best recommendations for implementation in Tables 4.1, 4.2, and 4.3 below for some examples). The 2020-2021 version of the Library of Congress recommended format standards now includes 3D data types²⁰. These file types are more interoperable and reusable and thus in line with FAIR principles. Human readable files formatted as ASCII text are easily parsed by both humans and computers for information regarding the documentation of a subject. Binary formatted data is an encoded form of the data requiring software to decode those data so they are human usable. Binary formatted files are generally smaller in size and run much faster when processed or executed by a program. However, decoding binary data is not assured deep in the future whereas human readable text is more likely to be parsed by humans many years from now.

As noted earlier, long-term preservation of 3D data is achieved by depositing open-format, non-proprietary files in a repository. However, proprietary formats of raw data offer indispensable metadata embedded in the file that are often lost in the process of converting the data to non-proprietary formats. In addition to interoperable open-source formats, there is a compelling argument to keep the data in its raw format as part of the local documentation of the project, if not the final archive. Work done by the Software Preservation Network (SPN) to ensure continued access to previous software versions and advances in **emulation**^{xxx} software increases the probability that researchers will be able to retain access to these important data structures. As appropriate, 3D data should be retained in their original file structure and format.

Good/Better/Best Recommendations for Implementation

Guidelines:

Good/Better/Best (GBB) recommendations are offered as guidelines to address:

- the level of documentation to target based on audience and use
- recommendations for file formats in consideration of access and/or preservation
- PIPs at which to save and preserve the 3D data
- databases or repositories that target the needs of the intended audience

This format of GBB can be found throughout the volume. Any level of 3D data preservation and documentation is a benefit to the community, and the Good level presented in this volume should be considered a target minimum for best practices. The tiered nature of preservation strategies is also seen in other frameworks, such as the Federal Agencies Digital Guidelines Initiative (FADGI)²¹ Guidelines and is not intended to place a value judgement on each level, but rather to create a structure that is easy to understand and remember. Not all 3D projects will require the highest level (Best) of preservation. Additionally, the level of preservation achievable will depend on the funding and resources allocated to the project. This is particularly true in regards to where the 3D data are preserved. At the time of this publication, many institutions do not have access to repositories with an infrastructure intended for 3D data. Project leads are encouraged to determine which tier

^{xxx} See Glossary

(Good/Better/Best) best fits their needs for preservation and access based on their intended audience, with a goal to get as close to that tier as planning and resources allow:

- **Good:** Preservation necessary for general access and basic use of the data. This tier targets general users, personal interest, and pre-packaged educational use of the final data package.
- **Better:** Preservation necessary for customizable educational use of the data and outreach. This tier targets reuse of the data for a variety of purposes, including both the final data package and the original raw data.
- **Best:** Preservation necessary for cultural heritage preservation and documentation of scientific and scholarly work. This tier targets the ability to reproduce and verify studies using the available information about the 3D data. Should include any proprietary data registration project files.

TABLE 4.1: GOOD: GENERAL USER, PERSONAL INTEREST, EDUCATORS, AND STUDENTS

	Recommendation	Examples
Level of Documentation	Context information about the data + Identify creation method/creator + Basic metadata information	Associated README text document with the recommended information Refer to see Chapter 4, <i>Metadata Requirements for 3D Data</i> , for examples of tiered implementation in metadata
Preserved File Formats	A program-neutral format	.obj, .stl, .ply, .fbx, .tif, etc.
Data Preservation Points	Final project files	
Long-term Access	Publicly accessible database	Thingiverse ²² , SketchFab ²³ , etc.

TABLE 4.2: BETTER: MUSEUMS, PUBLIC OUTREACH, NON-GOVERNMENT ORGANIZATIONS (NGOS), AND PROFESSIONALS

	Recommendation	Examples
Level of Documentation	Context information about the data + Identify creation method/creator + More robust metadata	Associated README text document with the recommended information + .csv files with specific metadata information Refer to see Chapter 4, <i>Metadata Requirements for 3D Data</i> , for examples of tiered implementation in metadata
Preserved File Formats	An open-source, program neutral format that is indexed by PRONOM	.obj, .ply, .stl, .tif, etc.

Data Preservation Points	Original “raw” data + Final Project Files	
Long-term Access	Publically accessible, institutional, or government repository	Dataverse ²⁴ , tDAR (The Digital Archaeological Record) ²⁵ , etc.

Table 4.3: BEST: SCHOLARS, RESEARCHERS, AND CULTURAL HERITAGE PRESERVATION

	Recommendation	Examples
Level of Documentation	Context information about the data + Identify creation method/creator + Robust metadata - including technical metadata ^{xxxi} outputs from software where available + Written explanation of the methodology (paradata)	Associated README text document with the recommended information + .csv files with specific metadata information + Technical reports associated with the software used (if available)
		Refer to see Chapter 4, <i>Metadata Requirements for 3D Data</i> , for examples of tiered implementation in metadata
Preserved File Formats	The original, proprietary format + An open-source, program neutral format, indexed by PRONOM, that allows for structured and customized embedded metadata	Proprietary: .fls, .skp, .vue, .psx, .txrm, etc. Open Source: .dae, .x3d, etc.
Preservation Points	Original “raw” data + Relevant Intermediary steps that preserve the decision making process + Final Project Files	
Long-Term Access	Institutional, government, or commercial repository with an infrastructure specifically for 3D data	MorphoSource ²⁶ , MorphoBank ²⁷ , Figshare ²⁸ , Nature Scientific Data ²⁹ , etc.

Conclusion

As the name implies, the CS3DP group is dedicated to creating community standards for the preservation of 3D data. The Preservation Best Practices chapter considers 3D as a new content type that must be brought into existing preservation best practices such as OAIS and TRAC. These criteria and considerations are well established for other content types; the challenge is to modify and customize them for the unique properties and workflows associated with the major 3D data collection and processing methodologies, while remaining flexible enough to accommodate new 3D data collection modalities.

^{xxxi} See Glossary

This chapter introduces the Preservation Intervention Points (PIPs) framework that details key assessment points throughout the lifecycle of a 3D project, and poses key questions to consider. An institution's priorities, capabilities, selection criteria, designated community requirements, and available funding will all inform how these questions are answered and what recommended actions are taken. The proposed framework can be applied across multiple 3D-data-generating modalities and tiers of implementation, from national to local governmental, educational, and private entities. As there can be no one solution that accommodates all institutions, all budgets, all data types, and all community needs, this chapter provides Good/Better/Best guidelines for baseline preservation recommendations that take into consideration these varying needs and available resources.

For those interested in seeing how the PIP framework and the Good/Better/Best guidelines work in a real-world context, please see Appendix X for six case studies that provide in-depth descriptions of several 3D data acquisition methods to demonstrate how to apply the recommendations from this chapter into existing workflows.

¹ Digital Preservation for Libraries, Archives and Museums, 2nd edition. 2017. E.M. Corrado & H. Moulaison Sandy, eds. Rowman & Littlefield.

² ISO 14721:2012 ([CCSDS 650.0-M-2](#)) Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model

³ ISO 16363:2012 ([CCSDS 652.0-R-1](#)) Space data and information transfer systems -- Audit and certification of trustworthy digital repositories

⁴ The PREMIS Data Dictionary for Preservation Metadata, <https://www.loc.gov/standards/premis/>

⁵ Wilkinson, M.D. et al. (2016). 'The FAIR Guiding Principles for scientific data management and stewardship'. Scientific Data 3:160018. DOI:10.1038/sdata.2016.18.

⁶ <https://www.go-fair.org/fair-principles/>

⁷ Archaeology Data Service (2008) Venus Preservation Handbook. <https://archaeologydataservice.ac.uk/research/venus.xhtml>

⁸ 'Data Selection: Preservation Intervention Points'. Archaeology Data Service / Digital Antiquity (2011) Guides to Good Practice. http://guides.archaeologydataservice.ac.uk/g2gp/ArchivalStrat_1-3

⁹ ISO 16363:2012 ([CCSDS 652.0-R-1](#)) Space data and information transfer systems -- Audit and certification of trustworthy digital repositories

¹⁰ <https://www.coretrustseal.org/>

¹¹ Witt, Michael and Carlson, Jake R. 2007. "Conducting a Data Interview." Libraries Research Publications. Paper 81 https://docs.lib.purdue.edu/lib_research/81/

¹² <https://www.doi.org/>

¹³ <https://orcid.org/>

¹⁴ <http://www.oais.info/>

¹⁵ <https://www.loc.gov/preservation/digital/formats/sustain/sustain.shtml>

¹⁶ DCC 'Five steps to decide what data to keep'. <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>

¹⁷ The Consultative Committee for Space Data Systems (2012) Reference Model for an Open Archival Information System (OAIS). <https://public.ccsds.org/pubs/650x0m2.pdf>

¹⁸ http://wiki.dpconline.org/index.php?title=4.2.1.4.2_Preservation_Description_Information

¹⁹ http://culturalheritageimaging.org/Technologies/Digital_Lab_Notebook/

²⁰ <https://www.loc.gov/preservation/resources/rfs/>

²¹ <http://www.digitizationguidelines.gov/>

²² <https://www.thingiverse.com/>

²³ <https://sketchfab.com>

²⁴ <https://dataverse.org/>

²⁵ <https://www.tdar.org>

-
- 26 <https://www.morphosource.org/>
 - 27 <https://morphobank.org/>
 - 28 <https://figshare.com/>
 - 29 <https://www.nature.com/sdata/>

Best Practices for Preservation – Appendix 1

The case studies below provide in-depth descriptions of several different 3D data acquisition methods to demonstrate how the PIP framework can be applied to existing workflows. These case studies are written from the perspective of a member of the project team. The purpose of each case study is to provide the necessary information to understand how the PIPs framework can be employed in different 3D imaging modalities. Because of the nature of the technology, in some instances terminology specific to the digital software or hardware for the project may be present.

Each case study examines a particular method of 3D data acquisition, the type of data collected, and the audience for those data to identify preservation and documentation needs for the project. Where it exists, accompanying text from a README file is provided as an example of best practices. At the end of each case study, a table identifying the Preservation Intervention Points (PIPs) for that project and recommended actions are provided for reference. The case studies are presented in alphabetical order - Born Digital, High-Resolution X-ray CT, Large-Scale Laser Scanning, Multi-Modal data collection, Photogrammetry, and Structured Light object scanning.

Each of the case studies presented here originate from different institutions, for different purposes, and illustrate diverse methods and technologies for curating 3D data. Recommendations for implementation following the “Good/Better/Best” (GBB) guidelines given in the main text are purely contextual based on an individual institution’s capabilities and resources for preservation. A single GBB assessment is given within each case study based on the method being illustrated and the respective institution to provide a practical explanation of how GBB is applied in the preservation workflow. These case studies are written from the perspective of the data creator and do not address SIPs and AIPS as expressed by the OAIS model for data preservation. For more information on repository operation, review the Long-Term Preservation section in this chapter or the Management chapter for more in-depth information.”

Born Digital

CASE STUDY: Lhasa VR 3D GIS and Virtual Environment
PROJECT DATES: 2009 - 2016
AUTHOR: Will Rourk, University of Virginia

Acquisition Method

Born Digital / Manually Created 3D
ESRI ArcGIS Desktop 10.7
ESRI CityEngine 2019.0
ESRI Unity 3D v5.6
Autodesk 3D Studio Max 2017
Agisoft PhotoScan 1.3

3D content was generated from the 2D GIS mapping of historic Lhasa and converted to an interactive virtual 3D environment using the Unity 3D game engine. 2D GIS is a method of creating maps with features linked directly to data either locally or from an online database. 3D GIS is a

method of converting 2D maps in an X,Y coordinate system to 3D content in an X,Y,Z coordinate system, allowing for spatial dimensions of geographical features to be emulated. Specific 3D models of historic buildings were manually created in 3D Studio Max, and 3D Data were collected from the photogrammetry of historic monuments using photographs from field work on site. These 3D assets were incorporated into the 3D GIS as detailed in supporting information. An online 3D Virtual Environment was created for interaction with 3D content through a web browser using the WebGL JavaScript API.

Good/Better/Best Assessment

The University of Virginia Library supports general methods of preserving data without specifically focusing on 3D data. A Best method of GBB is generally applied to data generated by 3D data collection methods or manual modeling. 3D assets associated with this project can be easily archived in the UVA Library's open scholarship platform, Libraⁱ (based on Harvard's Dataverseⁱⁱ platform), such as historically referenced 3d models created in 3D Studio Max or generated using photogrammetry. Archiving 3D GIS data is a procedure that is currently being researched due to the complexity of content involved. GIS projects involve many dependency files and file formats that are unified by a central project file. The UVA Library is moving in the direction of examining the use of data containers and emulation environments to archive the complexity of GIS projects and the myriad assets involved. Due to the current research conditions of archiving the complexity of GIS projects, a GBB assessment is difficult to administer. A "Best" method of archiving all project assets is achievable given the flexibility of the Dataverse open-access repository to ingest any type of uploaded file format. This is also true for the UVA Library's long-term archive, the Academic Preservation Trust.ⁱⁱⁱ

This project was part of a larger Mellon funded effort under the umbrella of the Humanities Virtual World Consortium (HVWC)^{iv}, in which the University of Virginia was a contributing member. The main goal of research of this group was to create a platform for creating and exploring digital narratives with 3D interactive technologies. The main software platform adopted for this task was the Unity 3D game engine^v augmented by Drupal for custom interaction within the 3D model/world.

The UVA team contributed to the HVWC by focusing upon Lhasa, the historic spiritual and political capital of Tibet. The objective was to create a comprehensive geographic information system (GIS) of historic Lhasa and then convert the map data to a 3D GIS that would then be converted into interactive 3D for the HVWC platform. At the core of this effort was the generation of a 3D model of Lhasa and the Kyichu River valley in which it resides. Two methods were used to create 3D content: manual 3D modeling in 3D Studio Max and procedural modeling based on GIS data in CityEngine. All phases of the 3D data curation centered around the creation of individual models that were incorporated into the larger city model.

Acquisition Process

Planning

Coordination happened at two levels: within the HVWC and inside the UVA team. The HVWC planning included choosing tools to build the interactive 3D platform and selection of third-party programmers to help build custom code. Specific functionality of the platform was a major planning issue because interaction with 3D content needed to match the criteria of the four contributing institutional partners.

At UVA a team was assembled that included Tibetan historians, GIS experts and a 3D content specialist. An historical narrative was created with a temporal focus on Lhasa as it grew from the mid-17th c. CE to early 20th c. CE. The team planned to first create a comprehensive GIS model of historic Lhasa, prior to Sino occupation in 1959. Many historic buildings were demolished and removed after 1959 so the intent of the GIS was to identify as many buildings as possible that once composed the city. Much of the planning centered around identifying maps, aerial images, text and other documentation that would reveal original position, use, and appearance of these buildings -- information that would be crucial to building the GIS of historic Lhasa. It was decided that the data would be held in the University of Virginia's Tibetan and Himalayan Library (THLib)^{vi} Places database, an existing repository of buildings, sites and place feature within the extents of Tibetan cultural areas.

PLANNING RECOMMENDATION:

This project involved many forms of data brought into a single platform for access and exploration. Organizing these data may require different techniques such as databases, GIS, and digital asset management tools.

Collection

Data collection began with obtaining documentation needed to build an accurate representation of pre-Sino-occupied Lhasa. This included the location of several historical maps, high-resolution USGS declassified aerial photos from the 1960s, SRTM digital elevation data and the selection of historical images of Lhasa from the early 20th c. CE purchased from the Pitt Rivers Tibetan image collection at Oxford University.

Architectural information was gathered from quintessential books of Tibetan architecture including *The Lhasa House: Typology of an Endangered Species*^{vii} and *The Temples of Lhasa*^{viii}, both by Andre Alexander, and *The Lhasa Atlas*, by Knud and Amund Sinding-Larsen. These books provided detailed plans, sections, elevations and textures to manually build key buildings in Lhasa by hand in 3D.

COLLECTION RECOMMENDATION:

Data collection is synonymous with research in this stage. Research asset collection methods can be best aided by consulting a librarian for strategies.

Processing

The processing stage was the most intensive for this project, as it was where the 3D content was generated using several methods centered around the construction of an interactive 3D GIS model:

1. GIS: A 2D GIS model was first constructed using ESRI ArcGIS. The first task in creating the 2D map of Lhasa was to import the high-resolution aerial image from 1966 and digitize building footprint polygons on top of it. Most of Lhasa's historic buildings were still intact at this point in time and their outlines were clearly discernible from the aerial image. Peter Aufschaiter's map of Lhasa from 1948 was also initially used because it defined over 860 buildings with place names that existed in the city core of Lhasa at that time. After defining a building footprint, the polygon feature was then connected to an entry in the Tibetan and Himalayan Library's Places database. Over 3500 features were defined and entered into the database. SRTM digital elevation data were also imported and georeferenced with the 2D map data. Images from the Pitt Rivers collection were connected to position markers in the map that indicated where the original pictures were taken. Other historical maps were overlaid and georeferenced with the map data. Roads, bodies of water, and vegetation were also mapped according to information from the historical maps and other documentation.
2. CityEngine 3D GIS: The 2D GIS data were imported via shape files into ESRI CityEngine, a tool for creating 3D models from 2D GIS data. 3D digital terrain was generated from the SRTM digital elevation data to create the landscape of the Kyichu River valley. Buildings were auto-generated based on the data from the THLib Places database. A height field existed in the database that indicated how many stories the building has in elevation. Code was written to take that field parameter and generate a place holder 3D model by extruding the 2D footprint to an elevation in the building height field. Buildings were textured with a generic image based on traditional Tibetan architecture. Thus, 3500 buildings could be generated quickly based on the backend data. Roads, rivers, and pathways were also auto-generated based on GIS features.
3. 3D models: Custom 3D models of key buildings in Lhasa were hand-modeled in 3D Studio Max. These were based on the architectural information from the books listed above. The models were not extremely detailed but showed more information than the auto-generated models. The completed model was then exported out of 3D Studio Max and imported into CityEngine and connected to the THLib places database via a unique identifier.
4. Unity 3D: Once the model was fully composed within CityEngine from procedurally-created and manually-constructed 3D models, the whole model needed to be transferred to the Unity 3D game engine platform. Elements were grouped together and exported simultaneously, such as buildings from different zones in the city, river features, road features, terrain, etc. These groups were then imported into Unity 3D.
5. HVWC platform integration: The Unity 3D interactive model was then integrated with the custom code from the HVWC, which was downloaded from the 3rd party developers via GitHub.

PROCESSING RECOMMENDATION:

In this case, processing is complex and highly project-specific. However, thorough documentation of this processing is essential.

Curation and Long-Term Access

It was decided by the Consortium that projects would be stored by the respective institutions.

The UVA project site is stored within the Tibetan and Himalayan Library as is the database backend. Project and construction files are backed up in cloud storage. Plans to make these assets available to the scholarly community are still being decided. The original code for the consortium is available on GitHub^{ix}.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

Archival platforms must be able to handle several files and their dependency structures.

Output

The main output for this project is the Unity 3D online model and Drupal code dependencies. Many files were generated in the construction of this model:

GIS data:

ArcGIS & CityEngine: .shp shape files, .tif geoTIFF (SRTM), .gdb geodatabase
WebGL ArcGIS webviewer

Database:

Tibetan and Himalayan Library Places Database - PostGres database

3D content:

3D Studio Max: .obj, .fbx, .dae
CityEngine: .fbx
Unity 3D: .unityproj, export to .fbx, WebGL

Web Content:

Drupal dependencies

OUTPUT RECOMMENDATION:

Complex projects generate a complexity of file types. Flexibility of storing and archiving diverse file types and maintaining their dependency directory structure is essential.

Usage

Usage is two-fold: use of the actual site and use of the HVWC platform. A website was created to show four different historic spatial narratives of Lhasa^x. The main narrative is the photographic account of Lhasa made by London Times special correspondent Perceval Landon in 1904. His tour around the historic city is retraced within an interactive 3D model with accompanying photos and

notes by the author. A guided tour is provided but an independent tour can also be undertaken. This resource is open for public use and exploration.

The intent of the HWVC was to release the code and a completed platform to the scholarly community for use with 3D content generated by other institutions. At 2nd phase funding, the project needs more work and refinement before it can be effectively released. The existing code is open for community access on GitHub as noted above.

USAGE RECOMMENDATION:

This is not quite a case of “once you have the data you can do anything”. But it is a case in which not only is the end product platform a useful tool, but so are the individual models and assets that went into creating the content for the platform. Making these available, as well as the end product, is recommended.

Born Digital Case Study: Preservation Intervention Points (PIPs)

PLANNING PIPS	ACTIONS
Build Partnerships	Divide responsibilities and agree on standard recording method.
Identify data sources and forms	Determine need for databases or digital asset management tools to organize information.
COLLECTION PIPS	ACTIONS
Research existing information	Utilize existing standards for research asset collection and consult a librarian for strategies.
PROCESSING PIPS	ACTIONS
GIS - annotate building footprint and add polygon feature	Define and enter features into the database (e.g., elevation data, images, historical maps, roads, water bodies, and vegetation).
Export GIS for CityEngine	Retain a copy of the original GIS file and the exported shape files.
CityEngine - additional data added	Record data added and location of source.
Add custom 3D models	Note which models were auto-generated and which were custom. Record source for decisions made during custom 3D creation
Unity 3D	Retain a copy of CityEngine files and exported files that were imported into Unity 3D. Maintain file structure for grouping together exports.

Platform Integration	Retain completed application. Document GitHub repository.
Curation and Long-Term Access PIPS	
	ACTIONS
Review files to determine preservation needs	<p>For this project, the following file types were preserved:</p> <p>GIS data:</p> <ul style="list-style-type: none"> • ArcGIS & CityEngine: .shp shape files, .tif geoTIFF (SRTM), .gdb geodatabase • WebGL ArcGIS webviewer <p>Database:</p> <ul style="list-style-type: none"> • Tibetan and Himalayan Library Places Database - PostGres database <p>3D content:</p> <ul style="list-style-type: none"> • 3D Studio Max: .obj, .fbx, .dae • CityEngine: .fbx • Unity 3D: .unityproj, export to .fbx, WebGL <p>Web Content:</p> <ul style="list-style-type: none"> • Drupal dependencies
Archive with each partner institution	Work with institution to maintain file structure - necessary for file dependency. Preserve assets as well as end product.

Case Study: High-Resolution X-ray CT (HRXCT) of Burmese Amber

CASE STUDY: High-Resolution X-ray CT of Burmese Amber
 PROJECT DATES: February-July 2019
 AUTHOR: Jessica Maisano, The University of Texas at Austin

Acquisition Method

High-Resolution X-ray Computed Tomography (HRXCT): uses polychromatic X-rays to nondestructively detect differences in materials within an object based on their density and atomic number. It produces a 3D digital map of the object's structure that can be digitally resliced, rendered in 3D, segmented into its constituent parts, analyzed quantitatively, and used to produce a surface model that can be rapid prototyped at any scale.

Good/Better/Best Assessment

This workflow exemplifies a 'Better' implementation of the tiered 'Good/Better/Best' approach to 3D data preservation. It exceeds 'Good' primarily because raw data are preserved, and these and the final data product(s) are preserved in perpetuity. It falls short of 'Best' because iterative versions

of the data (from raw, to client deliverables) are not preserved, and the data will not be placed in a publicly-accessible repository whereby they could be repurposed.

Object

Specimens are pieces of Burmese amber (Cretaceous) from various mines around Myanmar (e.g., Hukaung Valley, Kamti Mine, Angbar Mine). These pieces contain everything from crab claws and ammonites to bird wings and lizards. Most preserve only a mold of the original organism, but some do preserve bone/internal structure. Client is a commercial gem dealer, collaborating with an academic client of the University of Texas High-Resolution X-ray CT Facility (UTCT).

Acquisition Process

Planning

Client has examined each piece and described contents/region of interest to UTCT for scanning. Specimens are shipped or hand-carried to UTCT.

Scanning parameters depend on size of specimen/size and location of region of interest/scientific importance of specimen. Sometimes multiple specimens can be acquired in a single scan depending on their shape and density. Client confers with UTCT as to desired resolution, enabling UTCT to determine the appropriate scanning geometry (i.e., volumetric, in which 1000 or 2000 slices are acquired in a single rotation of the stage; or helical, in which the stage also translates on its vertical axis during acquisition to enable high-resolution imaging of high-aspect-ratio samples).

PLANNING RECOMMENDATION:

Client provides a unique identifier and locality information for each specimen, as well as an indication of the region of interest for scanning and desired resolution.

Collection

Most of these amber specimens are scanned on UTCT's North Star Imaging (NSI) scanner, although some requiring high-resolution close-ups are scanned on UTCT's Zeiss MicroXCT 400. Scanning parameters are determined at acquisition. For example, those amber pieces containing bone are rarer and thus more scientifically important, but often the presence of bone cannot be determined until digital radiographs are acquired of the specimen during scan setup. Data are then acquired, using acquisition quality standards corresponding to the scientific importance of the specimen while keeping within the client's budget (i.e., UTCT charges by the hour; in general, the longer the acquisition time, the better the data quality or signal-to-noise ratio). The resulting README

text, called 'contents.doc' at UTCT, accompanies the data to the client, to our archives, and to any repositories.

Below are contents.doc files (Client/Company name changed) representing the three possible imaging modalities for these amber specimens at UTCT: NSI volume acquisition; NSI helical acquisition; and Zeiss volume acquisition:

/Data From README/
NSI volume acquisition:

University of Texas High-Resolution X-ray CT Facility Archive

Smith:

Ref-29602: Scan of amber with insect (Ref-29602; Cretaceous, Myanmar) for Mr. Smith of Smith International. Specimen scanned by Matthew Colbert on 24 June 2019.

16bitTIFF: Scan parameters: NSI scanner. Fein Focus High Power source, 120 kV, 0.14 mA, no filter, Perkin Elmer detector, 0.25 pF gain, 1 fps, 1x1 binning, no flip, source to object 133.85 mm, source to detector 1316.703 mm, continuous CT scan, 2 frames averaged, 0 skip frames, 3000 projections, 5 gain calibrations, 0.762 mm calibration phantom, data range [-20.0, 130.0] (grayscale adjusted from NSI defaults), beam-hardening correction = 0.1. Post-reconstruction ring correction applied by Jessie Maisano using parameters oversample = 2, radial bin width = 21, sectors = 32, minimum arc length = 8, angular bin width = 9, angular screening factor = 4. Voxel size = 9.72 μm . Total slices = 1755.

8bitJPG: 8bit jpg version of the above images.

Specimen Photos: jpg photos of the specimen.

/End Data From README/

/Data From README/
NSI helical acquisition:

University of Texas High-Resolution X-ray CT Facility Archive

Smith:

Ref-31068: Scan of a lizard in amber (Ref-31068; 47.350 ct) for Mr. Smith of Smith International. Specimen scanned by Matthew Colbert on 25 June 2019.

16bitTIFF: Scan parameters: NSI scanner. Fein Focus High Power source, 130 kV, 0.24 mA, no filter, Perkin Elmer detector, 0.25 pF gain, 2 fps, 1x1 binning, no flip, source to object 133.85 mm, source to detector 1316.703 mm, helical continuous CT scan, vertical extent 48.6 mm, pitch 8.1 mm, 6 revolutions, 3 sets, helical sigma 0.0, no frames averaged, 0 skip frames, 15000 projections, 5 gain calibrations, 0.762 mm calibration phantom, data range [-10.0, 200.0] (grayscale adjusted from NSI defaults), beam-hardening correction = 0.1. Voxel size = 9.72 μm . Total slices = 4155.

8bitJPG: 8bit jpg version of the above images.

Specimen Photos: jpg photos of the specimen.

/End Data From README/

/Data From README/

Zeiss volume acquisition:

Smith:

Ref-30248A: Close-up scan of a fang in amber (Ref-30248; Cretaceous, Myanmar, Kampti Mine) for Mr. Smith of Smith International. Specimen scanned by Jessie Maisano on 17 May 2019.

scan parameters: Xradia. 4X objective, 80kV, 10W, 4s acquisition time, detector 42 mm, source -37 mm, XYZ [-1238, 42501, 784], camera bin 2, angles ± 180 , 1261 views, no filter, dithering. End reference (45 frames, each for 4s). Reconstructed with center shift -1.5, beam hardening 3, theta 0, byte scaling [-40, 1100], binning 1, recon filter smooth (kernel size = 0.5). Total slices = 944.

REF-30248A rcp: Xradia recipe with scan parameters.

16bit: 16bit TIFF images reconstructed by Xradia Reconstructor. Voxels are 3.15 microns.

8bitjpg: 8bit jpg version of the reconstructed images.

/End Data From README/

Acquisition parameters to consider documenting include: kV/W of the source (determines penetrating capability/flux; higher energy required for higher atomic number/denser samples, whereas lower energy preferable for delicate samples or those with small differences in attenuation between their constituent parts); filtering of the X-ray beam (to block the lower end of the energy spectrum, thereby minimizing beam hardening and ring artifacts); binning the detector to decrease acquisition time/resolution [detectors are 2000x2000 pixel, but can be binned once (to 1000x1000 pixel) or, in the case of the Zeiss, twice (to 500x500 pixels)]; distance between source, detector and specimen (determines resolution, or voxel size, of the resulting data set); number of radiographs to acquire (the greater the number, the greater the signal-to-noise ratio of the resulting data set – but also, the longer the reconstruction time); and time spent acquiring each radiograph (combination of time per view/averaging of views, again determines signal-to-noise ratio of resulting data).

Recording the scanning parameters also permits those with knowledge of HRXCT to interpret artifacts they see in the resulting data (e.g., if the README file indicates that no X-ray prefilter was used, and the data are beam-hardened, then the reader knows to prefilter the X-ray beam if scanning a similar specimen).

In addition to HRXCT scanning, UTCT takes digital photographs of biological and paleontological specimens, for two purposes. First, these photographs document the condition of the

specimen(s) when received. Second, rendering programs may load HRXCT slice stacks in opposite orders; in some cases this may result in a 3D rendering of the digital volume that is the mirror image of the actual specimen. Most fossils are strongly asymmetrical, so it is important to have a photographic record of the specimen to ensure that renderings are not mirrored.

COLLECTION RECOMMENDATION:

Record all sample mounting/scanning parameters necessary for someone to replicate the scan. The client may have multiple samples that require the same scanning conditions for the data to be directly comparable. Also, it is not unusual for the client to return months or even years later with new samples requiring the same scanning protocol. Photograph the specimen if it is asymmetrical, to check that 3D renderings are not mirrored.

Processing

Post-acquisition processing may include a beam-hardening correction, reorientation of the data, changing byte scaling, and a post-reconstruction ring correction, among other things. On UTCT's scanners, the first three are applied to the raw projection images using the proprietary scanner software.

A beam-hardening correction addresses the 'hardening' (increasing mean energy) of the polychromatic X-ray beam as it passes through the specimen; left uncorrected, this artifact will make the specimen appear artificially more attenuating peripherally than in the center in the reconstructed slices. It is best to prevent beam hardening by filtering the X-ray beam (using glass, calcium fluoride, aluminum, brass or steel of varying thicknesses), but most scanners have a polynomial correction built into their reconstruction software. It is important to record the correction applied, and it is possible to over-correct.

Reorientation of the data involves rotating the volume so that it is more orthogonal to the specimen, if necessary. This can be done in the scanner reconstruction software, or after data reconstruction in a program like ImageJ. Reorientation is important to note so the client can relate the data deliverable (a stack of 16bit TIFF slices) to the raw projection images.

Byte scaling refers to the scaling of the HRXCT data to fill the available 16bit grayscale space. 16bit data present a range of 2^{16} possible grayscale values from black (value 0) to white (value 65,535). If voxels (3D pixels) in the HRXCT data volume have a value of 0 or 65,535, this means that variation in those voxels has been discarded. Thus, it is recommended to byte scale to leave space at each end of the histogram of grayscale values. Data can always be rescaled to increase contrast post-reconstruction using a program like ImageJ.

The post-reconstruction ring correction is a program written in IDL to remove ring artifacts from reconstructed HRXCT slices. Ring artifacts are common in CT data, because any slight differences

between channels or pixels in the detector will manifest as rings in the reconstructed slices (except in systems like Zeiss, where the specimen stage shifts by a few microns on X, Y and Z from one projection to the next – ‘dithering’ – to prevent ring artifacts). UTCT’sRingFree post-reconstruction IDL ring correction program includes parameters such as bin width, number of sectors, minimum arc length, angular screening factor, and restriction of the grayscale range to process, to tailor results to samples of varying shapes and densities and artifact intensity. These parameters should be recorded because, if care is not taken, the ring correction can introduce new artifacts into the reconstructed data.

The documentation of some of these processing parameters is underlined in this READMEtext:

/Data From README/

NSI volume acquisition:

University of Texas High-Resolution X-ray CT Facility Archive

Smith:

Ref-29602: Scan of amber with insect (Ref-29602; Cretaceous, Myanmar) for Mr. Smith of Smith International. Specimen scanned by Matthew Colbert on 24 June 2019.

16bitTIFF: Scan parameters: NSI scanner. Fein Focus High Power source, 120 kV, 0.14 mA, no filter, Perkin Elmer detector, 0.25 pF gain, 1 fps, 1x1 binning, no flip, source to object 133.85 mm, source to detector 1316.703 mm, continuous CT scan, 2 frames averaged, 0 skip frames, 3000 projections, 5 gain calibrations, 0.762 mm calibration phantom, data range [-20.0, 130.0] (grayscale adjusted from NSI defaults), beam-hardening correction = 0.1. Post-reconstruction ring correction applied by Jessie Maisano using parameters oversample = 2, radial bin width = 21, sectors = 32, minimum arc length = 8, angular bin width = 9, angular screening factor = 4. Voxel size = 9.72 μm . Total slices = 1755.

8bitJPG: 8bit jpg version of the above images.

Specimen Photos: jpg photos of the specimen.

/End Data From README/

PROCESSING RECOMMENDATION:

As with recording data collection parameters, it is critical to record data processing/correction parameters so that the final result may be replicated. Also, it is possible to over-correct data, especially for beam-hardening and ring artifacts, so recording the parameters used enables the data consumer to determine whether this has occurred.

Curation and Long-Term Access

Curation at UTCT is achieved via a Microsoft Access database. The database is organized by project and records metadata like taxon name, museum accession number, specimen storage location, arrival/departure date, and locality information. The database also records scanning parameters, from which the READMEcontents.doc files accompanying each scan are directly exportable.

Long before the National Science Foundation required it, UTCT has had a robust long-term data management policy. The goal is to never *have to* re-scan a specimen – unless the scanning technology has improved so much that the data will show significant improvement (this technology turnover occurs every ~10 years). The cost to maintain HRXCT data and data products generated is covered in part by the ‘archiving’ fee charged to the client, based on the total gigabytes of a project. This means that if the client loses their data at any time in the future, they can come back to UTCT for them.

UTCT archives all raw HRXCT data and client deliverables (e.g., reconstructed 16bit .tif slice stacks, data analysis spreadsheets, data visualizations, etc.) to duplicate external hard drives; one is kept on-site, and the other is taken off-site. All client deliverables are maintained on redundant on-line servers at UTCT, and backed up to the off-site Texas Advanced Computing Center (TACC) at UT’s research campus, and to the cloud-based UTBox.

Long-term access is available on the internal UTCT network to all of these resources. Short-term access is provided to the client via virtual server FTP and/or UTBox. Occasionally for large projects, data are sent to the client via external hard drive.

Long-term access is also provided to a subset of biological and paleontological data sets via DigiMorph.org, the Digital Library of Morphology. This NSF-funded Digital Libraries Initiative project resides at UTCT. And, via a newly funded NSF grant called oUTCT, approximately 9 terabytes of HRXCT data representing ~1,500 fossil and Recent vertebrate taxa will be uploaded to MorphoSource.org where they will be more easily discoverable and repurposed. In addition, various UTCT data sets are reposted as supplemental information in journals upon publication.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

At a minimum, the deliverables to the client (at UTCT, the 16bit .tif and 8bit .jpg slice stacks, any derivative image processing or data analysis, and README contents.doc file) should be retained, for whatever duration is possible. These should be retained in multiple locations in case of hard drive/server failure. In the best case, all data generated during a project (raw scanner files, iterative versions as data are corrected, and those components already mentioned) should be retained.

The philosophy at UTCT is to keep the accessibility (and generally, expense) of a particular data type in line with the likelihood of revisiting it. For example, the raw scanner files are rarely revisited, so they are archived on inexpensive external hard drives. The client deliverables, on the other hand, are kept in duplicate on on-line servers, because they are revisited much more frequently.

Output

Output from HRXCT scanning at UTCT ranges from raw scanner files to visualizations to packaged data analyses. These file types will vary from facility to facility.

The raw files produced by the HRXCT scanner prior to 2008 were individual .raw sinogram images. With the addition of the Zeiss scanner in 2008 came the proprietary .txrm and .txm raw file formats, which are a collection of digital radiographs ('chest films'). With the upgrade of the original scanner by NSI in 2013 came raw projection images in .tif format with associated .nsi proprietary files.

The analysis/visualization programs employed at UTCT also produce their own proprietary file formats: .hx (Avizo); .vgl and .vgp (VGStudioMAX); .ORSSession and .ORLObject (Dragonfly); and.sav (Blob3D).

Data acquisition deliverables to the client include a README file (Microsoft Word document) and 16bit .tif and 8bit .jpg versions of the HRXCT slices. The READMEfile includes all of the relevant scanning parameters and data corrections applied to replicate them if necessary.

Data visualization deliverables to the client may include .tif or .bmp frames of animations, .avi, .mov and/or .mp4 movies, and/or surface models (.obj, .stl, etc.) to use in FEA analysis and/or for rapid prototyping.

Data analysis deliverables to the client may include .xlsx spreadsheets generated from various analytical packages (e.g., Blob3D, ImageJ).

List of file types: .avi, .hx, .jpg, .mov, .mp4, .nsi, .obj, .ORLObject, ORSSession, raw, .sav, .stl, .tif, .txrm, .vgl, .vgp, .xlsx.

OUTPUT RECOMMENDATION:

Not all file types have longevity. UTCT decided to deliver data to clients as 16bit TIFF slice stacks because the TIFF file format has been around since the mid-1980s and is readable by all major data rendering/analysis programs. That said, 16bit images cannot be handled properly or loaded at all by many programs, so it may be necessary to convert them to 8bit. This transforms the scaling of the data from 65536 to 256 possible grayscale values spanning black to white. 16bit data are much more information rich, but the user can throw information away by converting them to 8bit – the opposite is not true.

Usage

Most UTCT clients are academic researchers seeking information about the internal structure of their specimen(s) for scientific research and publication. A smaller percentage is commercial clients seeking to reverse-engineer their specimen and/or determine what defects it may have.

For the 'typical' academic client, usage will involve the publication of the HRXCT data set and/or visualizations/analysis derived from it as figures/supplementary information; these data should be made available so that scientific peers can critically evaluate them, just as any other data. For commercial clients, usage will involve the identification of defects within the sample so they can be corrected during the manufacturing process, or reverse engineering to produce a superior product; these data typically are not seen outside of the commercial entity, and often not archived by UTCT at the request of the client. In commercial cases involving historical objects (e.g., violins), usage may involve using the HRXCT data to identify unnecessary restoration that can be removed. In other commercial cases, the data may be used to prove/disprove the soundness of the scanned object in a legal setting. In most of these commercial cases, the data will not be seen outside of the commercial entity.

Secondary usage, by those other than the original client, is diverse. It may include K-12 and secondary educators who wish to incorporate HRXCT data into their lesson plans and labs, or use these data to rapid prototype models for their classroom. It may also include academic researchers seeking to repurpose data for their own research. Artists represent another class which seeks to repurpose HRXCT data, either for digital modeling or rapid prototyping. Finally, publishers and producers often inquire after images and animations derived from HRXCT data for books and documentaries.

USAGE RECOMMENDATION:

If at all possible, make data available for repurposing. This accomplishes a number of objectives: 1) minimizes specimen handling that would be required for multiple scans, which is especially important for precious natural history specimens; 2) extends the impact of the funds used to acquire the data, which often are provided by the taxpayer; and 3) allows the data to be used to address new/different questions outside the scope of the original research objective.

There are several options for making HRXCT data available for repurposing. For specimens scanned at UTCT there is DigiMorph.org – however, this library has not had active NSF funding since 2007, so it is necessary to charge the client on a cost-recovery basis to construct new DigiMorph pages. That said, existing DigiMorph pages represent data that have been cited in more than 300 scientific publications – most of which are examples of repurposing.

Other options include general data-sharing sites like FigShare, Dryad, MorphoBank, Nature Scientific Data, and Open Science Framework. Most recently, UTCT has decided to reposit approximately nine terabytes of data into MorphoSource (Duke University), a repository with the specific target of volumetric data.

Wherever one chooses to reposit data to make it available for discovery and repurposing, keep in mind any requirements regarding data ownership/copyright that the lending and/or scanning institution may have (see [Copyright Ch.](#)), as well as ensuring proper acknowledgment of the source(s) that funded the data acquisition.

X-ray CT: Preservation Intervention Points (PIPs)

PLANNING PIPS	ACTIONS
Retrieve information from client	Determine and record scanning objectives based on needs and budget of client. Record unique specimen identifier from client.
COLLECTION PIPS	ACTIONS
Scan specimen using parameters finalized at acquisition stage	Create a README text (contents.doc) that documents the scanning and processing parameters and the types of files generated, which is shared with the client.
PROCESSING PIPS	ACTIONS
Beam-hardening correction	Record the correction applied.
Reorientation of the data	Record reorientation of the data so it can be related back to the raw projection images.
Post-reconstruction ring correction	Record the correction applied.
Curation and Long-Term Access PIPS	ACTIONS
Curate at UTCT	Save relevant data into the in-house Microsoft Access database, including client information, specimen metadata, and scanning parameters.
UTCT file preservation policy	Archive raw HRXCT data and client deliverables (.tif slice stacks, data spreadsheets, data visualizations, etc.)
Client deliverables	Deliver via short term FTP access or UTBox. Where applicable, upload to DigiMorph or MorphoSource.

Large-Scale Laser Scanning

CASE STUDY: Large-Scale Scanning of University Hall, University of Virginia
PROJECT DATES: September 2018 - May 2019
AUTHOR: Will Rourk, University of Virginia

Good/Better/Best Assessment

As mentioned earlier (see the Born Digital case study) the UVA Library employs a “best” method of preservation for 3D data due to the flexibility of both open access and long-term preservation platforms in use. Dataverse and AP Trust, respectively, can both ingest a wide variety of data formats including esoteric file formats generated by 3D laser scanning technologies. This includes the preservation of proprietary 3D data file formats associated with the processing of raw 3D scanner data. The data from a FARO Focus 3D scanner is not a single file but a directory network of file dependencies that includes the 3D geometric and color data as well as scanner settings used during data acquisition governed by a single .fls formatted data file. The metadata for acquisition settings and conditions are only included in the project file and since these data will be crucial for reproduction of the data in the future, the UVA Library accepts this file format and its dependencies as primary data for archiving and data preservation despite the proprietary, binary nature of the data. The raw project file can be exported to other more exchangeable formats such as .ptx or .e57, which are included in the data preservation package.

The 3D laser scanning process yields several 3D data sets necessary to fully document all surfaces and dimensions of a site or subject. The raw data processing of these datasets involves the registration and precise placement of all individual datasets in spatial relationship to each other to create a single comprehensive and unified dataset. The UVA Library preserves all individual raw datasets in addition to the registered data in the project file. Derivatives of the raw data can be exported to provide more exchangeable data formats that are more easily implemented in other software through immediate open access as well as useful for long-term preservation such as human readable, ASCII text formatted files which include .pts, .obj, .ply and .x3d. A “best” assessment of recommended preservation is given here because the UVA Library has the capability of preserving esoteric as well as widely usable file formats for a single 3D documentation project.

Acquisition Method

Laser scanning - large scale
FARO Focus 3D model scanner (S120, X130)

Terrestrial 3D laser scanning technologies acquire data using laser light to capture the surface geometry and onboard color cameras to capture color data of the subject. Also known as LIDAR (Light Imaging Detection and Ranging), this method uses light bouncing off a surface to calculate the distance between the laser light source and the subject. The resulting data comprise a mass of points (commonly called a point cloud) that represents the geometric surface conditions of the subject as the time of documentation. Most 3D laser scanners are devices of metrology and are used for measuring the dimensions of a subject. In this case study, terrestrial 3D laser scanning is used because the laser scanning device functions on a tripod located on the ground.

Object

University Hall (U-Hall) was once the central indoor sports facility on grounds at the University of Virginia. Built in 1965, the design was based on a very large dome with a compression ring wrapped around concrete piers connected to concrete roof structure ribs. The building was replaced in 2006 and demolished in May 2019. The building was considered to be beyond restoration especially due to massive amounts of asbestos covering the arena roof, which accounts for 99% of the exposed roof structure. Ten months prior to demolition, a group was assembled to perform 3D documentation of the building. The documentation teams included robotic 3D scanning of the interior (Department of Computer Science and Engineering), aerial and terrestrial photogrammetry (UVAS Institute for Advanced Technology in the Humanities - IATH), 3D CAD modeling (Department of Architecture), architectural history (Department of Architecture) and 3D laser scanning of the interior and exterior (UVA Library Scholars' Lab). The 3D laser scanning captured the most data. The focus of interior documentation was to capture major public spaces including the arena, main hallways and passages and any ancillary spaces that defined the extents of the building interior. The focus of exterior documentation was to capture as much exterior surface as possible and rely on aerial photogrammetry for rooftop data.

Acquisition Process

Planning

The first stage began with conversations with the University Lead Conservator, which led to a wider conversation with the Office of the University Architect to determine what resources might be called upon within the University to perform a thorough 3D documentation of the site. It was determined that data collected could be managed by the Library and made available to the scholarly community via the Library open access repository. A team was formed with specialization in the areas described above. Documentation of the site needed to coincide with activity still taking place at U-Hall. The arena had been sealed off from the rest of the interior to prevent any potential hazards from the asbestos-covered ceiling. It was decided that the main documentation effort would take place during winter break after the 2018 Fall semester, but before asbestos abatement scheduled to begin in the early part of 2019.

A preliminary site visit made a couple of months before project initiation included a concurrent visit with all documentation team members. It was decided that the building was too complex to capture every single room given the project schedule. Architectural plans were prepared for a preliminary walkthrough of the site to determine which spaces were essential for documentation. These spaces included the main arena, the main hallways and passages, public meeting spaces and ancillary spaces that describe the full extent of the building including utility closets and key office spaces.

U-Hall was quite well known in the region as a major basketball venue, so the architectural historians decided that a narrative could be developed along the path the basketball team would travel from the locker room to arena. Thus the space where the basketball team lockers once resided

was included as an essential space, as was the former “green room” where performers once prepared for shows in the arena.

The team was allowed into the arena for approximately an hour to discuss strategies for documenting the expansive space -- roughly 300 ft in diameter and 100 ft in height. It was decided that 3D laser scanning would capture most of the data but would be supplemented by an initial autonomous robot scanner and terrestrial photogrammetry. Aerial photogrammetry was decided against inside the arena to prevent any disturbance of asbestos.

The teams planned to do a test scan of the arena to ensure that the technologies on hand were capable of efficiently documenting such an expansive space. For laser scanning, the architectural plans were edited to highlight areas to be documented and approximate scanning positions to determine how many scans would be needed. These scan positions helped to determine the amount of time it would take to actually capture the building data, which could then help determine how much time it would take to process the data for a completed set of documentation.

PLANNING RECOMMENDATION:

Make sure all parties involved understand the nature and intent of the 3D data. Consider client wishes for levels of access to data.

Use architectural site documentation to plan scanning positions. If such documentation is not available, create hand-drawn sketches of the spaces to be documented. These are part of the documentation record, supplemental to the data.

Collection

The equipment used for laser scanning U-Hall were two FARO Focus 3D X130 and one FARO Focus 3D S120 laser scanners. There were 2 stages for the documentation:

1. Preliminary test

Due to the expansive space, a few laser scans were taken in advance of the actual scanning to test the sensitivity of equipment and quality of data gathered. This testing was concurrent with the robotic scanner testing, whose team ended up being captured as data by the laser scanning test. This proved to be advantageous as they provided scale figures in the final data to show the immense scale of the arena. Two FARO Focus X130 scanners were used for the preliminary test.

The following procedure is followed whenever starting a new scan project:

1.1. Choose scanners and tripods that are most effective for the task at hand:

Tripods include Manfrotto MT055XPRO3, Gitzo carbon fiber tripod, Nedo survey grade tripod.

Two FARO Focus 3D X130 scanners were chosen for the preliminary test with Manfrotto MT055XPRO3 and Nedo survey grade tripods.

- 1.2. Project name: After scanners are booted up and mounted on tripods, a new project needs to be created in the settings on the scanner. The file naming protocol is in the form ProjectName_scanner#_YYYYMMDD, e.g. UHall_s2_20181224. Scans will be iterated starting with either 00 or 01.
 - 1.3. Profile: A scanning profile is chosen to set parameters for data capture. The default setting for an interior space that is greater than 10 meters was chosen for the arena. This is approximately a 15-minute scan that captures 44 million points of data plus color photographic texture.
 - 1.4. Scanner location: Choose locations that are optimal for collecting data that effectively document the site. In the preliminary test, data were recorded in key positions on the floor, first tier and top tier of the stadium seating. Eight scans recorded enough information to show the project managers the extent to which the space could be recorded. Scanner location is marked on the “scanplan” or architectural plan of the site. The naming convention for marking a scanner location is in the form scanner#-scan#, ex. S3-01. This number is consistent with the naming of each scan dataset that was input at the Project naming phase above. A data and location number is also included in the title of the scanplan, which will also help identify datasets during the processing phase.
 - 1.5. Level the scanner: Once the scanner is in its location and will not be moved it must be leveled. The FARO Focus scanners have a built- in inclinometer that is accessed from the Sensors pane of the Manage window. This sensor acts like a digital dual axis level so that the scanner is completely level prior to scanning. The scanner will not scan if it is too far out of level, and leveling it also helps the data processing stage.
 - 1.6. Prepare for scanning: Ensure that no people are in the way of the scanner and the surfaces to be recorded. Also clear out anything that might be obscuring surfaces. The most important features of a space are the areas that show the extents, like wall corners and tops and bottoms of walls, and any architectural details important to the historic narrative, like cornices, fireplace hearths, window and door surrounds.
 - 1.7. Start the scanner: Initiate scanner by pushing the green Scan button on the Focus 3D screen. Try not to place yourself in the scan or obstruct any surfaces needing to be documented.
 - 1.8. Finishing the scan: When the scan is finished, return to the home screen on the scanner and prepare for the next scan.
2. Actual documentation

A week was planned for complete capture of data. This week coincided with winter break to maximize building access. The laser scanning was split into interior and exterior space documentation stages. Interior spaces were prioritized according to access.

2.1. Interior

Arena: The team was given a small window of time, just a few days, to enter the arena space. The teams then coordinated schedules with each other so as to not get in each

other's way. Two scanners were used to capture data, set diametrically opposite and moved around the circular space. A third scanner was brought in to scan the press box and entryway. Due to the asbestos threat, a certified P100 grade respirator was worn while working in the arena space.

Hallways and other interior spaces: Scanners were set up on both levels of the building to capture the major hallways. A third scanner was used to capture stairwells and utility spaces.

2.2. Exterior

Two scanners were used to capture the exterior, placed at opposite sides and moving in sync around the building. A third scanner captured data at all major entrances. Spherical targets were placed at these positions to help with the registration of interior and exterior spaces.

Total data collection took over 40 hours of scanning time yielding over 100 individual scan datasets.

A final dataset was collected five days prior to the implosion of U-Hall, just before charges were set and nearly 85% of the building structure had been manually demolished. This provided a dataset of the raw structural system of the area and dome roof. Scanning time was nearly three hours resulting in 16 scan datasets that were added to the original set of data.

COLLECTION RECOMMENDATION:

Scanning positions should be optimal to capture as much surface data of the subject as possible for thorough documentation.

Data resolution is a balance between sufficient data collection for thorough documentation and time needed to perform data collection.

Always do full 360° scans to capture more data than needed unless pressed by time constraints. Extra data is better than missing data.

Processing

Once the scanning is finished all the scan data must be downloaded and processed. The main goal of the processing phase is the registration of individual datasets (scan readings) into one comprehensive dataset. FARO Scene(v. 2018) is the main software used for registration of data from the FARO Focus 3D scanners. The main objective of registration is to ensure that all individual datasets register as accurately as possible to maintain data integrity.

This stage has distinct steps:

1. Download the data: Each scanner records data onto an SD card, which must be downloaded onto the processing workstation.

2. Create a FARO Scene project: Create a project with naming convention SiteName_CollectionDate_scanproj, e.g., UHall_20181226_scanproj.
3. Import data into Scene: After all the data has been downloaded, the individual datasets can be dragged and dropped into the Scene project. Each dataset is not a single dataset but a directory of dependencies that includes the raw data. The topmost directory is considered the scanner data file and this is what is imported into Scene. The data are in the form of raw point clouds. . Color photo data are also included and automatically matched to the point cloud geometry.
4. Organize data in Scene: It can be difficult to process all of the data at once, so the datasets must be organized into what Scene refers to as “clusters”. These are special folders which allow processing of sets of data rather than the entire data volume, and they comprise scan datasets that were collected in relative spatial proximity to each other. For example, all the arena data would be put in one cluster, the 1st floor data in another, 2nd in another, and so forth. It is recommended to further subdivide data organization into subclusters for more efficient registration. For example, the arena may be divided into sections like the main floor, north, south, east and west quadrants and the press box. The more atomized the datasets into clusters, the easier it is for Scene to process and register them. The scanplan will help determine which scans are within proximity to each other, which is why . it is important to carefully record scanning positions during the collection stage.
5. Registration
 - 5.1. choosing an algorithm: Once the data are organized into clusters, registration can begin. This involves selecting a cluster and applying the Place Scans Operation. There are three algorithms that Scene provides for registering data sets:
 - 5.1.1. Target based - Use only if targets were used during scan collection stage. Targets include spheres, checkerboards, flat circular targets, etc. Identify targets in the scan data, name them so they are consistent between scans. Scene will then register datasets using at least three targets that are consistent between datasets
 - 5.1.2. Top View - Scene can register datasets without targets using this algorithm by “viewing” the data from above and finding spatial similarities. The result is usually a rough association requiring further registration.
 - 5.1.3. Cloud to cloud - This is another “targetless” registration method where Scene compares data points in the datasets to find similarities by which datasets can be joined together. This process is usually used to supplement the rough registration provided by the Top View algorithm.
 - 5.1.4. Correspondence View - Scene provides this method for manually moving datasets within proximity of each other. A view of all datasets in a cluster is provided and individual datasets can be selected, moved and rotated in place. This is usually used during the Top View registration phase when Scene has difficulty relating datasets on a rough draft level.
 - 5.2. Checking registration accuracy: After each Place Scan registration, Scene reports back numerical statistics on how closely the datasets registered within millimeters of each

other. Scene will green light registration if it is within about 5-6mm of accuracy. A consistently yellow or red light registration result indicates that Scene is having difficulty finding relationships between scan datasets. Manual transformation of datasets using Correspondence View is usually the solution to obtaining a green light result.

- 5.3. Iterative Registration: Datasets generally need multiple iterations of registration (usually Cloud to Cloud) to obtain the best results based on the registration report.
6. Color the data: Once the data have been accurately registered, the color imagery can be applied to the geometry, creating a more photorealistic view of the data. Sometimes data are more easily interpreted in grayscale, in which case it is not necessary to color the data. Data from the U-Hall project were colorized for photorealistic appearance.
7. Create Scan Point Clouds: This process is native to FARO Scene and is a way of optimizing the point cloud so it can be viewed in different ways including Virtual Reality or video screen capture.
8. Apply clipping box: Creating scan point clouds also allows for the use of a clipping plane to help frame specific views of the data. The clipping box is a virtual cube in Scene that hides any data that is outside the box.
9. Architectural views: The clipping box can be used to create architectural views of the data including plan, section, elevation, isometric and perspective views.
10. Rendering architectural views: Once the clipping box is implemented to provide architectural views, these can be saved as screen captures for characterizing the 3D dataset outside of the Scene program.
11. Exporting data derivatives: The final phase of processing involves exporting datasets that will be included in the data archive as well as for immediate use. The dataset can be exported in total, which is recommended for archive, or from the clipping box, which is more efficient for immediate use in other programs.

PROCESSING RECOMMENDATION:

Use processing methods that allow for checking the accuracy of registration between datasets.

Curation and Long-Term Access

3D data at the University of Virginia are currently archived using three different methods:

1. Immediate backup via cloud storage and physical hard drive: This method is for safe backup of all files, both active and archivable, and is primarily for individual access.
2. Dataverse: This is an immediate access solution. This is an open source, open access repository that allows all assets to be fully accessed by the public. Data that are published in

Dataverse are made automatically searchable and discoverable in Virgo, the UVA Library's catalog.

3. AP Trust: This is a dark archive into which only the project file and first flush derivatives are placed for long-term storage.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

Consider different levels of archive access on a spectrum from open access to dark archive.

Output

Exported file formats for point cloud data include:

- .Isproj - This is the FARO Scene native project file that retains all metadata and detailed information from the scanner and from processing.
- .PTS - This is an ascii text format that preserves the data in its most basic form as seven values, x,y,z,i,r,g,b. XYZ are 3D geometry values, RGB are color values, and i describes the way light interacts with the surface.
- .PTX - This is a binary format that is a directory of file dependencies that includes photographic information as well as geometry.
- .PLY - This format can be ascii and/or binary and retains all color information with geometry. ASCII is preferred for archiving purposes, whereas binary is preferred for immediate use.
- .E57 - This is currently used as an open exchange format for point cloud data, although it is a binary file. It would be included for short-term archives.

OUTPUT RECOMMENDATION:

Data formats should reflect immediate as well as long-term data usage. ASCII formats are predicted to be more useful in the future because content is human readable and easily parsed. Propriety raw data formats will probably not be as useful in the future due to the need for specific software to interpret binary data, which will most likely be deprecated or dead in the near- to long-term future. However these raw data files, usually project files, contain most of the metadata and detail information that are lost when exported to other formats. In a best case scenario the project file should be archived to preserve metadata from the original capture of the data.

Usage

- 3D print
- Virtual Reality
- 3D modeling, CAD
- BIM
- GIS

USAGE RECOMMENDATION:

The data are the most important part of the process. Anything can be done once the data are responsibly collected and processed.

Large Scale Laser Scanning: Preservation Intervention Points

PLANNING PIPS	ACTIONS
Build team of experts to document the site	Determine responsibilities of each team member. Document this information and include required levels of access to the data by the client.
Conduct preliminary site visit	Walk through the site with architectural plans and determine which spaces would be captured with which methods. Retain this documentation as supplemental data.
COLLECTION PIPS	ACTIONS
Conduct test scans	Document equipment, parameters, and procedure used during test scans. Use test scans to determine extent of scanning and time required. Retain any documentation of agreements between parties.
Conduct documentation level scans	In addition to the information above, document scanning positions and data resolution. Ideally, this level of documentation will include additional data beyond the minimum needed for the project. Retain original scans as raw data.
PROCESSING PIPS	ACTIONS
Import data for registration	Retain a copy of the scene file into which original scans are imported.
Organize the data into clusters	Save iterations of working files as needed. If useful, record clusters used for registration process.
Register clusters	Create a protocol that accounts for the registration algorithm and iterative registration in a consistent manner or documents deviations from that protocol at a reasonable level. Retain a copy of the registered data and any additional processing applied (such as coloration, scan point clouds, or clipping boxes).
Architectural views & rendering	Create and render views as needed. Retain a copy of the rendered views and screen captures.
Derivative data	Export the data for archiving and further use. Archival preservation should include the total data regardless of a clipping box.
Curation and Long-Term Access PIPS	ACTIONS

Archive using University of Virginia standards

Immediate backup via cloud storage and physical hard drive, Dataverse, and AP trust.

Multi-Modal

CASE STUDY: Warm Springs Bath Houses 3D Site Documentation
PROJECT DATES: April 2016
AUTHOR: Will Rourk, University of Virginia

Good/Better/Best Assessment

Because laser scanning was used in this project, the explanation for a “best” method of data preservation is already given above (see Large-Scale Laser Scanning). The project illustrated here adds the photographic source material, processing and 3D content results from aerial photogrammetry methods using a quadcopter and high-resolution camera. In this case, the individual captured source photos intended for archive are the .jpg files from the quadcopter camera. The photogrammetry project built using Agisoft PhotoScan (v 1.3), resulting point cloud, and color texture mesh 3D content are also included in the archive. 3D content exported to similar formats as the laser scanning, such as .obj, .ply and .x3d, are included. The laser scanning data and photogrammetry data were combined using Reality Capture (v 1.0), by Capturing Reality^{xi}, to produce a comprehensive set of data from both types of 3D data. The Reality Capture project file and exported 3D data formats are also intended for preservation. As mentioned above (see Born Digital), the project files are important for reproducibility of data in the long term because they capture all data acquisition as well as data processing metadata. Due to the proprietary nature of the project file formats, it is understood that these files will be difficult to access over time as software becomes deprecated. It is hoped that data preservation methods, such as data containers and emulation environments being researched by the Software Preservation Network^{xii}, will increase accessibility to the information in project files over the long term.

Acquisition Method

- Terrestrial laser scanning
- Aerial photogrammetry
- Ambient audio recording

Terrestrial laser scanning is a form of LIDAR technology that uses light reflected on the subject surface to determine surface geometry and dimension as well as color imaging to capture surface texture (see Large-Scale Laser Scanning Case Study for more details). Data are acquired from a ground location. Aerial photogrammetry uses 2D images acquired from a UAS (Unmanned Aerial System) that are processed to create 3D data with color photo-realistic texture (see Photogrammetry Case Study for more details). Ambient Audio Recording is achieved with a device that records audio in 2-channel stereo or more channels to capture ambient sounds occurring at an existing location. Multi-modal is a technique of combining one or more documentation technologies (modes) for a comprehensive spatial record of the subject.

Object

The Jefferson Pools, also known as the Warm Springs Bath Houses, are located in the western part of Virginia, close to the West Virginia border. Two structures exist on the site, each housing a pool of sulfuric water fed by natural springs. The northern structure (for men) has an estimated origin in the late 18th century CE while the larger southern structure (for women) was built

in the 1870s. . In 2016 an assessment of their condition was performed due to decay and structural deterioration. It was requested that a full site 3D documentation be performed to help with the assessment and provide for the historical record. A team from the University of Virginia was assembled to acquire data using terrestrial 3D laser scanning and aerial photogrammetry with a quadcopter UAS. Because water is an important factor for the bathhouses, ambient audio was also recorded at key points where water interacts with the site.

Acquisition Process

Planning

The UVA team coordinated with a member of UVA Facilities Management who was part of the preservation efforts at the Jefferson Pools. The team also worked with architect Terry Ammons of StudioAmmons in Petersburg, VA who was charged with site investigation and generation of a Historical Structures Report (HSR) which served as the record of historical and present day site conditions as well as recommendations for preservation. Equipment chosen for data collection included two FARO Focus 3D X130 and one FARO Focus 3D S120 laser scanners for terrestrial scanning and a DJI Inspire quadcopter for the aerial data. Equipment preparations for the 3D laser scanners included cleaning the scanner optics, printing checkerboard targets, charging batteries and making sure the tripod was secured. Aerial photogrammetry preparation included cleaning the HD camera optics on the quadcopter, charging enough batteries for at least 2 hours of flight and making sure to pack extra quadcopter parts, like extra rotors and a battery charger as well as making sure the iPod tablet for the camera feed was charged and flight software updated.

The team travelled to the site where they worked with maintenance crews who maintain the site for the Homestead Resort Hotel in nearby Hot Springs, VA, owned by Omni Hotels. Upon arrival the team did a preview of the site to note key features to be documented and potential impediments such as locked rooms, unstable structural areas, water hazards and tourist access. Timing of aerial and terrestrial documentation processes was key to ensuring that one team did not impede the other. The decision was made to perform terrestrial scanning inside the structures while the aerial team worked outside to gather images of the entire site.

Initial site preparation included the placement of checkerboard targets throughout the interior and exterior of the site for laser scanning registration. Checkerboards were also placed facing upward on the terrain to coordinate with the quadcopter data collection. Green tennis balls were staked throughout to also be used as targets for aerial data. Site plans were drawn of each building as well as the overall site to record scanning positions.

PLANNING RECOMMENDATION:

In cases where the data acquisition process includes multiple teams, document expectations for each group including the timeframe when the work will be done. Account for the need for shared space and resources so that multiple groups don't need access to the same resources at the same time.

Collection

Aerial data were collected by a 2-person team, one navigating the quadcopter and one operating the HD camera on a 3-axis gimbal. Each task required operating individual control systems. To capture the entire site, the quadcopter was flown in rows across the site to capture terrain data, then flights were made encircling the structures at varying levels in accordance with photogrammetry methods. Single shot aerial photos were taken at higher altitudes restricted by FAA rules on flying UAS devices. Terrestrial photos were also taken using the quadcopter camera for views from the ground.

Laser scanning involved positioning three scanners around the site. Initial data collection began inside each bathhouse, with scanners placed in key positions to capture the most effective spatial data in each location. Major spaces were documented, but some minor rooms were excluded due to time constraints. Scanning positions were noted on the plans for each building with a naming convention of “scanner number - scan number”. For example, 2-3 next to a square locator symbol on the plan indicated the position of scan #3 from scanner #2. Recording scan positions is key to successful registration of scan data in the processing stage.

Audio was recorded at the site after the 3D documentation stage. Recordings were taken at key points in the site exterior where water interacts with the structures and flows through the site. Recordings were acquired using a Zoom H1 stereo audio recorder.

COLLECTION RECOMMENDATION:

Treat individual collection methods as normal. Maintain clear and documented file structures for the data that accounts for the different acquisition methods. Each set of data should be a self contained set of material that can be cross referenced with the additional data sets.

Processing

After the data were collected by the two different techniques, each set of data was processed independently and then combined:

1. Aerial Photogrammetry:
 - 1.1. Download digital photos from the DJI Inspire 1.0 quadcopter SD card.
 - 1.2. Cull photos for bad shots, e.g., blurry photos, photos out of scope, misfired photos, etc.
 - 1.3. Collect photos into directories for each major site feature, e.g. women’s bathhouse, men’s bathhouse, full site.
 - 1.4. Main software for photogrammetry processing was Agisoft PhotoScan Pro on a 12 node computing grid. The PhotoScan process is a very detailed process; the following is a basic outline of the processing procedure:
 - 1.4.1. Create a new PhotoScan project for each major site feature
 - 1.4.2. Drag and drop quadcopter photos into the PhotoScan Workspace
 - 1.4.3. Create sparse point cloud
 - 1.4.4. Optimize cameras and refine point cloud
 - 1.4.5. Create dense point cloud
 - 1.4.6. Generate color texture

1.4.7. Generate 3D mesh content

2. Terrestrial Laser Scanning:

- 2.1. Download scan datasets from each scanners' SD card.
- 2.2. Main software for registration of datasets is FARO Scene (v. 6.2) (for detailed workflow see the Large-Scale Laser Scanning Case Study).
- 2.3. After registration, color texturing and point cloud generation, export .ptx file.

3. Multi-modal Processing

Main software for processing laser and photogrammetry data is Reality Capture created by CapturingReality. The basic process for combining 3D laser with photogrammetry data is:

- 3.1. Import laser data .ptx file generated by registration in FARO Scene.
- 3.2. Import photogrammetry data from Agisoft PhotoScan.
- 3.3. Generate comprehensive 3D dataset coordinated by establishing control points between the disparate datasets.

4. Audio Processing

- 4.1. Audio files were downloaded from the Zoom recorder's micro SD card and stored on a local drive.
- 4.2. The main processing software was Audacity (v 2.1.2).
- 4.3. Audio files were clipped appropriately to create looping files in a 3D program.
- 4.4. Files were exported to .mp3 for use in 3D software such as Unity 3D (v. 5.6).

PROCESSING RECOMMENDATION:

Develop a protocol for each method individually as well as a protocol on how to combine the materials.

Curation and Access

3D data at the University of Virginia are currently archived using three different methods, which are described in this section of the Large-Scale Laser Scanning Case Study.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

Multi-modal data curation should account for file dependencies and maintain a file structure that clearly marks how the datasets integrate with each other.

Output

Exported data files are contingent upon the process implemented as well as 3D formats that are best suited for near term and future use:

1. Aerial photogrammetry:
 - a. Photos: .rawformat .dng, high resolution .jpg
 - b. 3D mesh file formats: .obj, .ply, .dae, .x3d
 - c. Project file: .pscan PhotoScan project file
2. 3D Laser Scanning:
 - a. 3D point cloud file formats: .ptx, .pts, .e57, .ply, .x3d
 - b. FARO Scene project file: .lsproj
3. Multi-modal Exports from RealityCapture
 - a. 3D Mesh formats: .obj, .ply, .dae, .x3d
 - b. RealityCapture project file: .rc
4. Sound recordings
 - a. .wav, .aiff, .mp3

OUTPUT RECOMMENDATION:

Consider a variety of data formats that address the variety of data types included in the project.

Usage

- Preservation: The point cloud dataset was used by Richmond architecture firm 3North to help their design work on mediating the structural integrity and maintaining the historic construction and design of the Jefferson Baths during the preservation reconstruction process.
- A VR model of the Jefferson Baths was created in the Unity 3D game engine and made openly available in the UVA Clemons Library public VR space. The virtual model of the Jefferson Baths was used as a multi-participant meeting space for the CHIVR (Cultural Heritage Informatics VR) initiative in the Spring of 2018 between UVA, James Madison University and the University of Mary Washington. The model was enhanced by including 3D audio objects based on the sound recordings. Sound was associated at key points in the site and could be treated as 3D objects in Unity 3D with spatial characteristics such as attenuation and sphere of influence. Files were set to loop for continuous ambient sound.

USAGE RECOMMENDATION:

Multi-modal projects have a variety of assets that can be re-used and re-mixed for a variety of purposes. Careful consideration of possible re-use scenarios can assist in determining the level of preservation.

Multi-modal Site Documentation: Preservation Intervention Points (PIPs)

PLANNING PIPS	ACTIONS
Choose technology specialists appropriate for site documentation needs	Determine responsibilities of each team member.
Schedule site access with project coordinators	Secure permissions from all groups involved.
Preliminary site preview	Walk through the site with architectural plans and determine which spaces would be captured with which methods. Retain this documentation as supplemental data.
Team coordination	Decide what areas need to be scanned when and by whom so as not to impede each other's progress or potentially block data acquisition.
Place targets for registration and control points	When needed, place targets in areas that will overlap between scanning locations for registration and highly visible areas for control points.
COLLECTION PIPS	ACTIONS
Conduct test scans	Document equipment, parameters, and procedure used during test scans. Use test scans to determine extent of scanning and time required. Retain any documentation of agreements between parties.
Conduct documentation level scans	In addition to the information above, document scanning positions and data resolution. Ideally, this level of documentation will include additional data beyond the minimum needed for the project. Retain original scans as raw data.
Laws and regulations	Be mindful of laws and regulations regarding the technologies employed for documentation, e.g., FAA rules on flying UASs
Record scan positions	Selection of key scanning locations should be backed up with proper documentation.
Supplement data with media	Audio and video recordings can supplement data acquisition and become helpful assets in the final archive.
PROCESSING PIPS	ACTIONS
Import data for registration	Document equipment, parameters, and procedure used during test scans. Use test scans to determine extent of scanning and time required. Retain any documentation of agreements between parties.
Vet content to be used for creating data	In addition to the information above, document scanning positions and data resolution. Ideally, this level of documentation will include additional data beyond the minimum needed for the project. Retain original scans as raw data.
Registration organization	Be mindful of laws and regulations regarding the technologies employed for documentation, e.g., FAA rules on flying UASs

Control points	Selection of key scanning locations should be backed up with proper documentation.
Derivative data	Audio and video recordings can supplement data acquisition and become helpful assets in the final archive.
Curation and Long-Term Access PIPS	ACTIONS
Archive using University of Virginia standards	Immediate backup via cloud storage and physical hard drive, Dataverse, and AP trust.

Photogrammetry

CASE STUDY: Photogrammetry of Artifacts from the Santa Cruz Museum of Art and History

PROJECT DATES: 2016-2018

3D DATA CREATOR: Abigail Crawford

AUTHOR: Marcia McIntosh, University of North Texas

Good/Better/Best Assessment

This workflow demonstrates a Good implementation of level of documentation, data preservation points, and long-term access, and a Better implementation of preserved file formats.

Acquisition Method

Photogrammetry is “the science of making reliable measurements by the use of photographs and especially aerial photographs”.^{xiii} It is a nondestructive and affordable option for creating 3D data. Freelance photogrammetrist Abigail Crawford^{xiv} completed the below case study.

Object

Almost 200 artifacts^{xv} that are housed at the Santa Cruz Museum of Art and History^{xvi}, including but not limited to a Doctor’s bag^{xvii}, Screaming Hand Sculpture^{xviii}, and Basket^{xix}.

Acquisition Process

Planning

Crawford consults with the museum to find out the purpose of the models (educational/outreach), in what format they want the final product(s), and how they would like the files to be named. She also gains permission from museum staff to do the photography, and, in some cases, contacts the artists to make them aware of the model’s creation and let them know they can use the data if so desired.

PLANNING RECOMMENDATION:

Match the intent for the model with the creation process. Intent can be based on the audience. Knowing the data’s intended purpose will affect the manner in which it is documented, the final outputs, and where the data will go for preservation.

Collection^{xx}

Camera: Rule of thumb: the higher the resolution, the better the photos, and as a result, the 3D model. A 5 megapixel (MP) camera or above is recommended (an iPhone 8 has a 7MP front-facing camera and a 12MP rear-facing camera.^{xxi}). Crawford started with a 5MP camera, then switched to a 24MP camera in April 2016 with a fixed focal length lens (50 mm).

Other equipment: A tripod and a remote control for the camera to decrease blurring of the images, and a turntable for the objects being photographed.

Setup: Place targets (paper with text or lines) under the subject to help align photos in the processing phase. The targets are not moved during the shooting session to avoid photo alignment issues later. A circular polarizing filter (CPF) is recommended when shooting shiny/reflective objects to reduce bright spots, although this may not work with metallic surfaces. If shooting outdoors, overcast/cloudy days are preferable. If shooting inside, use diffuse 360-degree lighting when possible. It was not always available at the museum.

Subjects: Avoid subjects that are untextured, completely flat, very thin, transparent, shiny, or reflective. Select items that are solid, matte, and textured.

Capture: Shots are taken in .raw format. Depending on the subject, images are captured every 10-15 degrees horizontally and vertically with roughly 50-60% overlap between photos. Before shooting, the subject is studied to see what areas or features may require more images for an accurate model beyond the standard rotations. It is recommended to take more photos, particularly when the object cannot be revisited, then select those that will be used in processing.

COLLECTION RECOMMENDATION:

Organize files according to institutional/departmental source, object, and status (raw, final, etc.).

Document the setup (camera, degrees per shot, etc.), and equipment settings, in a README.txt, .csv, spreadsheet, or similar file, to the extent that would be useful for the intended audience. Scholars/Researchers will want this information to accompany data submitted to a publication and submission to a repository.

Processing

If processing in Agisoft PhotoScan, the images are converted to .tif in Adobe Lightroom or Photoshop. If processing in Reality Capture, they are left in .raw. The files are organized based on the naming convention given by the museum. Unless asked to use accession numbers, descriptive names are given to the data.

- (folder) Santa Cruz MAH >
 - (folder) 4-11-18 >
 - (folders) Red Hat, Boardwalk Clown, Projector >
 - (files) projector1.dng, projector2.dng, etc.

- Final model format: Projector.obj, Projector.jpg, Projector.mtl
- (folder) Rosicrucian Museum >
 - (folders) Hand_cross_68, Shabti_79, etc. >
 - (files) shabti_79_1.arw, shabti_79_2.arw, etc.
 - Final model format: Shabti_79.obj, Shabti_79.jpg, Shabti_79.mtl
- (folder) Stand Up >
 - (files) stand_up_1.arw, stand_up_2.arw, stand_up_3.arw, etc.
 - Final model format: Stand_Up_250k.obj/mtl/jpg, Stand_Up_1M.obj/mtl/jpg

The final format for the model is .obj. The mesh may be edited in Blender, which can import the .obj file and export either it as either a .obj(for PhotoScan) or .dae(for Reality Capture) file.

Each step in processing is periodically saved to prevent loss. Blurry or unfocused photos are removed so they will not cause alignment problems. Depending on the demands of the software, the images may go through the time-intensive process of masking before alignment. The mesh is then created. Once it is satisfactory, it is simplified to reduce the size to the number of polygons requested by the museum. The mesh is then imported into Blender as a .obj file. There, extraneous geometry is removed, and holes and other geometry problems are fixed. The mesh is next exported from Blender and imported as a .dae file if using Reality Capture or as a .obj file if using PhotoScan. It is unwrapped and textured. After texturing, it is examined to make sure it looks accurate and exported into a .zip file. It is then sent to the clients through Dropbox. A pre-decimated mesh is kept in the project file.

PROCESSING RECOMMENDATION:

All photogrammetry data creators will want to save periodically during the processing steps to prevent data loss from software crashing or mistakes.

Cultural heritage preservation and scientific workflows should additionally document what processing choices are made for additional transparency and to assist in recreation of their models. This documentation may be accomplished by exporting a processing report from the photogrammetry software.

Curation and Long-Term Access

The files are put on an external hard drive along with the documentation and correspondence from the museum. The files are uploaded to the organization’s Sketchfab pages^{xxii}. The files are scheduled to be reviewed for data management and culling every ~10 years.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

At the least, have an internal backup of the raw data and final data model. It is recommended to make the data available online if possible. If used for cultural heritage or scholarly purposes, data products should be submitted to a repository.

Output

The clients receive: .obj, .mtl, texture images (These can be .jpg, .tif, or .png depending on the situation), and, optionally, the .raw photos.

OUTPUT RECOMMENDATION:

Depending on the intent of the model, .obj, .mtl, texture images, README file, processing report, and other descriptive metadata may result in a completed photogrammetry dataset.

Usage

The intended usage for the Santa Cruz Museum models is outreach and educational purposes.

USAGE RECOMMENDATION:

The data can be consumed by the intended audience or others once accessible online.

Photogrammetry: Preservation Intervention Points (PIPs)

PLANNING PIPS	ACTIONS
Consultation with Stakeholders	Determine audience and use to guide documentation and file format needs. Acquire rights and permissions as needed.
COLLECTION PIPS	ACTIONS
Setup space	Record use of targets and equipment to create a better environment or shot (such as circular polarizing filter or additional lighting)
Capture images	Save .raw images. Record camera settings and interval of photos.
PROCESSING PIPS	ACTIONS

Option 1: Process images for Agisoft PhotoScan	Organize files using given naming convention. Save captured .raw files as .tif.
--	---

<ul style="list-style-type: none"> ● Process model in Agisoft PhotoScan 	Save working files periodically. Record mesh reduction options. Export Agisoft metadata file. Preserve final PhotoScan file.
--	--

<ul style="list-style-type: none"> ● Blender processing 	Retain copy of .obj exported from PhotoScan. Record if extraneous data was removed, if holes were filled, or other types of processing done.
--	--

<ul style="list-style-type: none"> ● Finalize model 	Retain copy of textured export from PhotoScan.
--	--

Option 2: Reality Capture Processing	Use .raw photos for processing.
--------------------------------------	---------------------------------

<ul style="list-style-type: none"> ● Process model in Reality Capture 	Process and save periodically in Reality capture.
--	---

<ul style="list-style-type: none"> ● Blender processing 	Export and use .obj in Blender to fix holes or extraneous geometry. Export from Blender as .dae and retain a copy of this file.
--	---

<ul style="list-style-type: none"> ● Finalize model 	Retain a copy of the textured export from Reality Capture.
--	--

Curation and Long-Term Access PIPS	ACTIONS
---	----------------

Personal archiving	Save relevant files from previous stages with documentation, including correspondence from the museum.
--------------------	--

Institutional archiving	The clients receive: .obj, .mtl, Texture Images (, .tif, or .png depending on the situation), and, optionally, the .raw photos as per original agreements.
-------------------------	--

Online public access	Upload derivatives to organizational Sketchfab pages as agreed.
----------------------	---

Create a review period	Review retention needs every 10 years.
------------------------	--

Structured Light Scanning

CASE STUDY: Lithic Collection from the Early Upper Paleolithic Site of Tvarožná X, Czech Republic

PROJECT DATES: 2016-2019

AUTHOR: Kristy Golubiewski-Davis, University of California Santa Cruz

Acquisition Method

Structured Light Scanning (SLS) via a single camera David SLS Scanner, now Hewlett-Packard (HP) SLS Scanner

Structured Light Scanning is achieved by projecting a series of known patterns onto an object using white light. The software measures the distortion of the pattern from a known point which is calibrated at the beginning of the process. These measurements are used to create a point cloud which, in the case of HP Scan software, is interpreted into a polygon mesh for processing and export.

Good/Better/Best Assessment

This project exemplifies a 'Better' implementation of the tiered 'Good/Better/Best' approach to 3D data preservation. The data are archived in the UMN institutional repository and will also be archived in tDAR for long-term storage. The choice to follow 'Better' was intentional and reflected limitations in available data storage. Raw data were preserved, but neither the intermediate data nor the proprietary files associated with the project were preserved in perpetuity. A copy of these files is accessible through a departmental share drive as an intermediate-term storage solution. This drive is backed up nightly, but is not a long-term archival solution.

Object

This project was part of an effort to create 3D documentation of the lithic artifact collection from the Early Upper Paleolithic open-air site of Tvarožná-Za školou (Tvarožná X) to supplement traditional excavation methods^{xxiii}. The 3D dataset^{xxiv} represents the 187 retouched tools, complete flakes, and cores that were over 2 cm in length.

Acquisition Process

Planning

The primary concern in the planning stage was to determine the optimal parameters for the scans without having the lithic artifacts present, given that the artifacts were in the Czech Republic and the researchers conducting the scanning were in the United States. Two methods of 3D data

acquisition were considered: photogrammetry (see description in Photogrammetry Case Study) and structured light scanning. The research team opted to bring both sets of equipment to the scanning location, and spent the first day testing.

Prior to going to the field, the research group decided on the desired parameters for scanning. These included: how to determine the selection criteria for scanning (Tools, Cores, and Refits that are greater than 3cm along the widest access); which scanning methods would be tested for use (structured light and photogrammetry); and the resolution, number of scan perspectives, and rotation degree of the turntable that would be used for each method. The 3D data would be acquired in the field and processed at a later date in the lab, necessitating the collection of additional data to compensate for scanning errors. A repository was identified (the institutional repository + tDAR), but not consulted at this stage. Consulting a tDAR representative at this time would have benefited the project, as it affected the final derivative file format and the method through which tDAR collects project metadata.

PLANNING RECOMMENDATION:

When planning to acquire 3D data of material that is unavailable for testing and has unique characteristics, build in time for on-site testing. As possible, pilot test scans prior to travel and determine the functional requirements of the project and associated scanning parameters. Build in time beyond the expected time for the procedures to account for scanning issues. Discuss the project with any repository resources early to build their limitations and requirements into the workflow, especially if the repository does not have an existing workflow specific to 3D data deposition.

Collection

After testing, it was decided that the best scanning option given staff expertise and time was the HP SLS scanner. A scanning protocol was created during the planning stage and refined during on-site testing. The parameters of the protocol are listed in the README file associated with the data, and are included below for reference. As scanned, information about each artifact was entered into a Google Form, including any deviations from the protocol. For the scanning stage, the following information was captured:

- Unique Identifier (associated with the artifact's accession number)
- Times: start time, end time, total scan time
- Additional scans taken to capture the edge
- Total number of scans captured
 - Additional notes as needed
 - Additional perspectives taken
 - Issues during the scan
 - Known refit and conjoin information

/Data from README/

The artifact was scanned on a David SLS 2 scanner using the 30mm calibration plate and the David automatic turntable.

The object was lit using the room lighting.

The artifact was placed vertically and flipped for 2 perspectives unless otherwise stated in the metadata file.

For each perspective, the artifact was automatically turned every 20 degrees 18 times for a complete 360 degrees and scanned at each angle.

If needed to capture the edge, the flake was manually turned so that the camera viewed the edge directly. A scan was taken and the edge was turned a maximum of 5 degrees twice in either direction to capture a total of 5 scans per edge.

/End Data From README/

COLLECTION RECOMMENDATION:

Determine a set of known parameters that will be the baseline for 3D data capture. Use a consistent method of recording any deviations from those parameters. Google forms worked well for this project, as it was easy to use in the moment, created consistent data, and provided a spreadsheet output which was repurposed for the structured metadata that is currently associated with the dataset.

Processing

All of the 3D data were scanned first and processed at a later date due to time restrictions at the artifact location and computer processing power. As with collection, all of the artifacts went through the same protocol in the processing stage with deviations collected via a Google form. The protocol was made available alongside the dataset in the repository, and can be found below. Multiple save points were kept during the working phase, with four points in the workflow identified as points to collect data via google form and ensure a save copy for internal preservation. Each form included a series of hand edits, which were recorded as present or absent for each artifact. Definitions for those edits were provided in the README file to provide transparency on the editing process. The following data was collected via Google Form during those points:

- Unique Identifier (to link the data from other forms)
- Times: start time, end time, total scan time
- Number of alignment attempts
- Number of perspectives
- Vertex spacing - after merge and final
- Merge resolution
- Hole filling - present / absent and program used
- If hand editing was performed in the following manners:
 - Complete mesh removal

- Edge deletion
- Edge duplication
- Original edge reintroduces
- Partial mesh removal
- Picked point alignment
- Texture errors removed
- Labels
- Seams
- Mesh defects
- Defeatured (Geomagic X)
- Smoothed (Meshmixer)
- Flattened (Meshmixer)

/Data From README/

Initial alignment of the scans was completed in David 4 software using the automatic align feature.

Additional cleanup, global alignment, and hard to align scans were completed in Geomagic Design X. Specific processes used during this phase are noted in the Pre-Merge Data Editing Processes column of the file TvaroznaX_3D_metadata.CSV.

The photo textures were processed using Lightroom 5.5 and Photoshop CS6 to better approximate the original texture. This processing included using auto tone and auto balance to brighten the image, removing the background, and refilling the background with a color based on the midrange color of the artifact.

The final merge was completed in HP 3D Scan Pro (formerly David 5). Scans were set to merge with an estimated vertex spacing of .075 mm or less. The merged result was exported as an .obj and imported into Geomagic Design X.

Next, scan texture was converted to vertex color. The scans were centered around the origin, and oriented based on standard conventions for lithic analysis.

The mesh was copied, and manual editing operations were performed as noted in the Post-Merge Data Editing Processes column of the metadata sheet. If Meshmixer was used, the scan was exported from Geomagic, the required operation were performed, then the new mesh was reimported.

A global remesh operation was run with the target average edge length set to .075 mm. Next, the healing wizard operation was run to mitigate mesh defects. Color information was copied from the original mesh imported into Geomagic.

The mesh was then exported from Geomagic as an ASCII PLY file.

/End Data From README/

PROCESSING RECOMMENDATION:

Identify points in the process to retain working copies of the 3D data. The HP SLS scanner creates shelled data. The working copies for this project included four consistent processing save

points, which effectively acted as PIPs: first-round edits just before merge; final edits just before merge; just after merge; and final object. Collect documentation on deviations from the protocol in a manner consistent with the collection stage.

Curation and Long-Term Access

Three preservation strategies were followed for this project. Each strategy reflects different use case needs for the data and limitations of the strategies available at the time of publishing the dataset. The metadata and paradata collected with the project were collected into both a README text file and structured .csv file that are associated with the data in each instance. The use of Google Forms early in the process expedited the process of creating both the structured .csv data, which collected the most relevant information, and the unstructured README text file. The data in tDAR and DRUM include only the final derivative data and neither the raw data nor the working files, reflecting the funding available for preservation and preservation strategies discussed with both repositories. The differences in file format reflect the requirements of the repositories. The original derivative data were in .ply format and were exported again to .obj for tDAR.

The three preservation strategies followed for this project were:

1. Maintain a copy of identified preservation working points, raw data, and final derivatives on departmental drives. This allows the researchers to share the raw data if requested and the working files that are most likely to be used if reprocessing of the data is required. This is not a long-term preservation strategy, but the research team determined it was the best option available for the full set of data associated with the project.
2. Preserve a copy of the final derivative data in .ply format via DRUM (Data Repository for U of M), the institutional repository of the PI. Documentation is provided at the project level via a README file and a .csv file.
3. Preserve a copy of the final derivative data in .obj format via tDAR (currently in process). Documentation will be provided at both the project level, in a manner similar to the DRUM repository, and at the object level as README files.

CURATION AND LONG-TERM ACCESS RECOMMENDATION:

Discuss preservation strategies with the intended repository early, including file format options. Have multiple strategies and consider options for preserving raw data and mid-process data that captures important workflow decisions.

Output

Derivatives for the project include .ply files and .obj files. Both file types include textured images. Proprietary file formats include the original scan files (.davidproj) and the Geomagic .x3d files that were created during the processing stage.

OUTPUT RECOMMENDATION:

Ensure that any raw data or derivatives for the project are available in both proprietary and non-proprietary formats.

Usage

The intended usage for the scans is for scholarly and research purposes. Specifically, the models can be used to view identifying information such as the bulb of percussion and ripples and provide models detailed enough to allow for meaningful measurements to be taken from the 3D data.

USAGE RECOMMENDATION:

Ideally, the preservation and public accessibility stage of this project would have included the raw files. Though not possible for this case, the raw data are being kept on networked departmental drives that are backed up regularly. Additionally, detailed notes on the 3D creation and processing stages were kept to document the work done to the data.

Structured Light Scanning: Preservation Intervention Points (PIPs)

PLANNING PIPS	ACTIONS
Determine functional needs of the project	Record expected parameters and protocols. Develop a method to consistently acquire data.
Consider space planning needs	SLS requires control over lighting conditions and access to power. Incorporate possible issues into the expected timeline and protocols.
COLLECTION PIPS	ACTIONS
Scan capture	Record information that deviates from protocol or expected parameters. Recommended information to record with SLS: number of scans; number of scan orientations; and object data not already collected elsewhere that affects the scan. Retain a copy of the raw files in both proprietary and non-proprietary formats.
PROCESSING PIPS	ACTIONS

First round pre-merge processing	The first round of processing should stick as close to the protocols as possible. Record any variables and retain a working copy at this stage. If the initial test merge is good, move to texture processing.
Second round pre-merge processing	The second round of processing should include additional hand edits required to achieve a satisfactory merge. Record additional work as needed. Retain a working copy of the data at this stage.
Texture processing	If texture editing is part of the protocol, retain a copy of the textures both before and after any image processing.
Merge	Retain a working copy of the data just after the merge point in the workflow. Document any merge settings used that aren't already documented in the protocol.
Post-merge processing	Document any post-merge processing done to the data. Retain a copy of the final data in both proprietary and non-proprietary formats.

Curation and Long-Term Access PIPS	ACTIONS
Internal preservation strategy	Determine an appropriate location to store raw, working, and access-level data. Clean up the file directory prior to closing out the project.
Public access preservation strategy	Long-term preservation in publicly accessible locations should include documentation from earlier stages in a structured format accompanied by a document describing the structure of the data. Ensure data is accessible via both proprietary and non-proprietary file formats.

ⁱ <https://www.library.virginia.edu/libra/>

ⁱⁱ <https://dataverse.org/about>

ⁱⁱⁱ <http://aptrust.org/>

^{iv} <http://virtualworlds.etc.ucla.edu/>

^v <https://unity.com/>

^{vi} <http://www.thlib.org/>

^{vii} Alexander, André. *The Lhasa House: Typology of an Endangered Species*. Serindia Publications, 2019.

^{viii} Alexander, André. *The Temples of Lhasa: Tibetan Buddhist architecture from the 7th to the 21st centuries*.

Chicago: Serindia Publications, 2005.

^{ix} <https://github.com/HVWC/HVWC-Multiplayer-Platform>

^x <http://bit.ly/LhasaVR>

^{xi} <https://www.capturingreality.com/>

^{xii} <https://www.softwarepreservationnetwork.org/>

^{xiii} Merriam-Webster Dictionary

^{xiv} <http://www.archaeologicalgraphics.com/photogrammetry>

^{xv} <https://sketchfab.com/santacruzmah/models>

^{xvi} <https://santacruzmah.org/>

^{xvii} <https://sketchfab.com/3d-models/doctors-bag-4846fe7d8fd4499ca1e62ea24b895b09>

^{xviii} <https://sketchfab.com/3d-models/screaming-hand-15f218f80cde4286929d3b9c4c66107d>

^{xix} <https://sketchfab.com/3d-models/basket-dc3d8c07769e4226833cbd01a2936939>

^{xx} Crawford, Abigail. "How to Set-up a Successful Photogrammetry Project" (2015)

<https://sketchfab.com/blogs/community/how-to-set-up-a-successful-photogrammetry-project/>

^{xxi} iPhone 8 specifications. <https://www.apple.com/iphone-8/specs/>

^{xxii} <https://sketchfab.com/santacruzmah/models>

^{xxiii} Škrdla, P., G. Tostevin, D. Nývlt, L. Lisá, O. Mlejnek, A. Přichystal, D. Richter, 2009. "Tvarožná-Za Školou. The Results of 2008 Excavation Season." *Přehled výzkumů* 50(1-2):11-24.

^{xxiv} Tostevin, Gilbert; Škrdla, Petr; Monnier, Gilliane; Golubiewski-Davis, Kristina; Porter, Samantha, T. (2019). Lithic Collection from the Early Upper Paleolithic Site of Tvarožná X, Czech Republic. Retrieved from the Data Repository for the University of Minnesota, <https://doi.org/10.13020/hmrj-a558>.