community standards for 3D data preservation

Presentation for IASSIST 2018
Jennifer Moore, j.moore@wustl.edu, Washington University in St. Louis
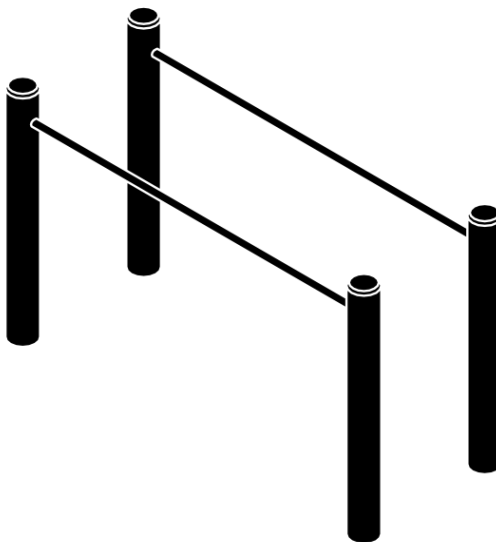Adam Rountrey arountre@umich.edu, University of Michigan
Hannah Skates Kettler hannah-s-kettler@iowa.edu, University of Iowa

INSTITUTE *of* **Museum** and **Library** SERVICES

3D Data Boom

- Equipment improved

- Models produced are better

Data Curation Priorities

- Push for organizations to prioritize data curation as part of the mission

Created by Lluisa Iborra
from Noun Project

Stewardship of scholarly record = established responsibility of libraries & museums

Data curation community actively makes a case for curation of research data & develops practices

Many of which apply to 3D data, but this datatype poses unique problems
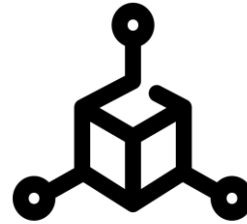
Who Cares About 3D Preservation?

Jennifer Moore
Washington University in St. Louis

Hannah Scates Kettler
University of Iowa

□ pdf

INFORMATION

Searching for Guidance

- London Charter, high-level principles for the preservation of 3D data

- The Guides to Good Practice from the Archaeological Data Service (ADS), basic suggestions for digital 3D data archiving in Archaeology

- 3D-Icons, which funded a report for the European Commission on 3D metadata to expand the Carare Standard

- Curating Research Data, brief case study describing a 3D preservation workflow

Growing list of resources:
https://drive.google.com/open?id=1s4s80XdQ13ZNFCGMiAPAVSnejjraiOFejZZjU1WtixM

High-level principles from the London Charter

PRINCIPLE 1- IMPLEMENTATION

The principles of the London Charter are valid wherever computer-based visualisation is applied to the research or dissemination of cultural heritage.

PRINCIPLE 2 - AIMS AND METHODS

A computer-based visualisation method should normally be used only when it is the most appropriate available method for that purpose.

PRINCIPLE 3 - RESEARCH SOURCES

In order to ensure the intellectual integrity of computer-based visualisation methods and outcomes, relevant research sources should be identified and evaluated in a structured and documented way.

PRINCIPLE 4 - DOCUMENTATION

Sufficient information should be documented and disseminated to allow computer-based visualisation methods and outcomes to be understood and evaluated in relation to the contexts and purposes for which they are deployed.

PRINCIPLE 5 - SUSTAINABILITY

Strategies should be planned and implemented to ensure the long-term sustainability of cultural heritage-related computer-based visualisation outcomes and documentation, in order to avoid loss of this growing part of human intellectual, social, economic and cultural heritage.

PRINCIPLE 6 - ACCESS

The creation and dissemination of computer-based visualisation should be planned in such a way as to ensure that maximum possible benefits are achieved for the study, understanding, interpretation, preservation and management of cultural heritage.

Guides to Good Practice Common 3D Formats Matrix

Format	Properties / Technologies	Description	Recommendations
.x3d	ISO standard XML-based format developed by the Web3D consortium	The X3D[1] format (extensible 3D graphics) was developed by the Web3D Consortium and has been ISO certified since 2006. The format is suitable for the storage of single 3D models as well as of complex 3D content such as virtual reality. The format succeeds VRML and, as such, should be preferred to it. The format is core to 3D in HTML5.	Suitable for preservation and recommended for complex 3D content.
.dae	COLLADA[2] XML-based exchange format	COLLADA (collaborative design activity) is an XML-based format developed by the not-for-profit Khronos Group consortium and designed as an interchange format for complex 3D data. ISO/PAS 17506:2012 provides a standardised specification for the COLLADA schema[3].	Suitable for preservation and recommended for 3D content where x3d is not an option.
.obj (also includes optional .mtl and .jpg files)	Wavefront OBJ file	The open OBJ format was developed by Wavefront Technologies and is supported by a large user community with open specifications[4]. The format stores both geometry and textures and consists of an obj file (ascii or binary format) together with an mtl (material/texture) file and image (actual texture).	Suitable for preservation of wire frame or textured models. ASCII format is preferred for preservation.
.ply	Stanford polygon file format	The PLY format, also known as the Stanford Triangle Format, is a simple format with ASCII and binary versions developed at Stanford University primarily for 3D scanning data. The format is inspired by OBJ but allows extension to incorporate a variety of properties including colour and transparency, surface normals, texture coordinates and data confidence values. The format also allows different properties for the front and reverse of a polygon. While the addition of certain extensions will not make the format unreadable, not all software supports all extensions so data may at best be unread or, in a worst case scenario, be discarded when files are resaved.	Suitable for preservation (ASCII version) although file content should be clearly documented.
.vrml	Virtual Reality Modelling Language[5]	Virtual Reality Modelling Language is a text-based standard (ISO/IEC 14772) for representing 3D interactive vector graphics and is the predecessor of X3D. The most recent version was published in 1997 as VRML97	Suitable for preservation although now replaced by X3D.

http://guides.archaeologydataservice.ac.uk/g2gp/3d_2-3.

3D Icons: key metadata principles

Provenance: technical information (equipment, light, barriers, software, processing)
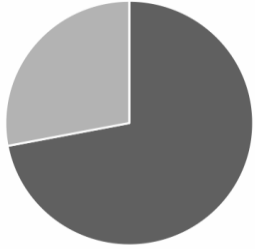
Paradata: human information (evidence used, methods, context)

Interim report on 3D metadata: https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/3D-ICONS/Deliverables/3D-ICONS_D4.2%20Interim%20Report_Metadata_Creation.pdf

-Andrea D'Andrea

Survey of the Community

Not Using Standards

- 72% of all respondents do not use best practices or standards

- Of this group, 69% said it was because they were unaware of such standards

Collaboratively Build Standards

- Those who were using standards largely developed them in-house

- 85% of all respondents said they would like to collaboratively develop standards & best practices as a community

https://osf.io/ewt2h/wiki/home/

Lack of guidance, and without guidance…

- Institutions are often scrambling to fill the gaps with ad-hoc, localized solutions

    (….or in-attention)

- Local solutions can create barriers to effective data stewardship & sharing

- Local solutions may inhibit efforts for large-scale, national or international data aggregation

Year of 3D Data

INSTITUTE *of*
Museum and **Library**
SERVICES

- Community Standards for 3D Data Preservation

- Developing Library Strategy for 3D & VR Collection Development & Re-Use

- Building for Tomorrow: Collaborative Development of Sustainable Infrastructure for Architectural & Design Documentation

CS3DP Plan

Cultivate investment in the development of standards & best practices that:

- the community can build or agree on together, instead of continuing on alone

- work to make our agreed upon standards responsive to the needs of all community members

CS3DP Plan

Pull together Community of Practice made up of

- researchers

- librarians

- data curators

- repository managers

CS3DP Outcomes

- Community-developed plan

- Recommendations for standards and best practices

- A report and publication

Which aim to make digital 3D data

Sharable - Discoverable - Accessible - Reusable

for the long-term

CS3DP Forum 1 - Experts from the US & abroad

Based on survey responses planned panels & discussions & working group meetings to develop:

- Preservation Best Practices

- Management and Storage

- Metadata

- Copyright and Ownership

- Discoverability and Access

Used town hall meeting to set the course.

Presentation Panel on Preservation best practices

Data Curation Best Practices for Data Preservation - Johnston

tDAR & 3D Scan Data - McManamon

Preserving 3D data: best practices from a UK perspective - Nivon

Guiding Questions	Discussion Points	Direction for Working Groups
How do existing practices for digital preservation apply to digital 3D data? What are current digital curation practices and how do they translate to 3D data? What guidance exists or is being developed to provide steps for forward migration and format longevity of 3D data? What is raw 3D data and how do we record digital provenance? How can we work around proprietary file formats?	Define what is a 3D model; it's more than just a given output. Identification of authentic, raw data is undetermined. Important to establish sustainable funding model for long-term curation and storage. Define what we mean by long-term. Calculating costs requires consideration of human time. Develop sense of user needs and how to assess them. Archivists should be involved in the selection and context of data. Need to be able to identify risks via a presentation plan. Preserved models need to somehow reflect the creators intent.	Guides for creators (incl. rights) on what to keep (and not) and what is needed (incl. obligations/contracts). Guides for what to share. Conduct assessment of costs. Assess how modality relates in these guides. Find a way to capture processes, workflows and use cases. Assess what are the limits of current preservation. Develop priority levels for preservation. File format guidance for communities. Capture info lost in conversion. Identify alternative methods for archiving complex interactive models.

MorphoSource: helping determine & serve 3D community needs - Boyer

What data should be kept over the long term?	The use of subscription models for contributors to repositories varies.	Scan existing platforms.
How does one track digital provenance?	Subscription models should not be at the user level.	Development of platforms.
Should we track digital provenance?	The role of augmented data undefined; what if augmented and resubmitted.	Identify resources/tech requirements, human time/skills.
What constitutes a master/archival copy?	DOI essential for citation, reproducibility and preservation.	Solutions for migration.
What options are being evaluated for storage of large data sets "in perpetuity"?	Understanding multi-model workflows and how they link together is essential.	Identify institutional repositories; create registry; centralize?
Cost/benefit analysis for storage solutions?		Explore problem of immediate open access vs. dark archives.
What challenges are repositories facing with this data type?		Explore solutions for ingestion/ technical validation.

Presentation Panel on Metadata Standards

3D Models for Cultural Heritage: Survey to shared knowledge - d'Andrea

Building Consensus: Overview of Metadata Standards Devo - Harlow

3D Metadata at Indiana University - Hardesty

Building metadata for 3D capture workflows - Blundell

Why do we need metadata standards development? How are standards developed? Given the lack of agreed upon standards for 3D data, what solutions are institutions currently using? What are users' needs regarding metadata? Who are the targeted users?	Configuration of declarative workflows needed. Important to look at the metadata model established by 3D Icons for the CARARE metadata. Questions of how metadata models needs to be. Need to make use of existing resources; not reinvent the wheel. Establishing metadata and data vocabularies essential because terms are defined differently based on domain. Our metadata model must consider born digital 3D data as well as capture. Essential to build metadata model that facilitates interoperability and reusability. Schema is good for capturing data about object, and not a specific technology, to grow with field. Descriptions of creation should be procedural as well as technical.	Define scope of 3D that are included (e.g. PG, laser, CT/volume). What is needed for minimal element set. What is transparency? Creation data - workflows vs. detailed metadata. Review existing standards and explore mapping to existing.

Presentation Panel on Metadata Standards

3D Models for Cultural Heritage: Survey to shared knowledge - d'Andrea

Building Consensus: Overview of Metadata Standards Devo - Harlow

3D Metadata at Indiana University- Hardesty

Building metadata for 3D capture workflows - Blundell

What constitutes the minimal metadata for inclusion in a repository? How can linked metadata be developed for improved workflows. What would be required for data clean up or migration for previously unstandardized data?	Flat metadata vs. semantic metadata. Need it to be readable by people and machines. Metadata needs to also describe algorithmically generated data and data that was interpolated rather than captured. Metadata needs to describe provenance of physical to digital object to be used for comparative analysis. High-level guidelines will help bridge the gap between treatments of different modalities (e.g., surface vs. CT scans). 3D models may be subjective and may be produced with a some uncertainty; lichard scale may help describe confidence level. Use of paradata may best support the existence of subjectivity in 3D models	Decide flat vs semantic metadata. Create a user guide on metadata / tools. Find tools for extracting embedded metadata. Find standardized descriptions by LOD.

Presentation Panel on Copyright/Ownership
US-based & International issues regarding 3D - Courtney & Levine

What is the current interpretation of copyright law to 3D Data in the U.S. and abroad, and is it appropriate? What can be learned from case studies of copyright from other media? Who owns the data? Who owns copyright on collaborative project data? What are strategies for negotiating agreements for content with limited rights due to permits or cultural sensitivity?	Data is not subject to copyright, so often times it's a matter of contracts and licensing. Many of the questions around 3D data remain untested in law. Open data is essential because commercially protected data would make it harder to openly publish. We can set the tone now on how we want our policy to look.	Develop user guide - when copyright when contract/license. Evaluate from raw to derivative chain. Develop tools to understand when it's creative/original. Find and provide sample workflows / process to evaluate and list rights/choices. Guidance on sharing sensitive material/public. Understand ethical restrictions/consultation- not just "legal".

Presentation Panel on Discoverability/Access

Sharing and Publishing 3D Data: Enabling Access and Reuse for Research Applications - Wittenberg

Preserving Multimodal Sources for 3D Digital Editions & Intersectional Scholarship - Nieves

Using Open source platform to improve Cultural Preservation - Mbunzama

What platforms are being used to share 3D data?	Government agencies don't consistently support making public domain data available.	Create a list and review of available repositories.
What challenges are repositories facing?	Considerations of data that exists in government agencies may have baggage due to past wrongdoings. Sensitivity needs to be figured into this problem. NAGPRA (Native American Graves Protection and Repatriation Act) influential in making sure people are held accountable.	Guide for user to identify reliable model.
In what state are users expecting 3D access?		Criteria for peer review.
How is the lack of consistent metadata standards impacting the ability to share 3D data?	Preservation is a social justice issue, but it's important that we don't just add thin layers of social justice to 3D preservation; this means involvement of affected communities.	Guidance on data citation.
How can linked metadata impact discoverability?	We need to maintain the linkage between physical objects (e.g., at museums) and models to preserve both.	Explore 3D reviewers list.
Who are the targeted users?	May need to establish a model of peer-review.	"Watermarking".
		Assess screen reader.
		Certifications/NINES -Nineteenth Century Scholarship Online.

Moving forward…

- Working groups are meeting virtually on their own schedule and have tasks and goals

- Have developed tools to further survey 3D practices
  https://osf.io/ewt2h/wiki/home/

- Full group is meeting virtually, monthly

- Working groups will present findings at next forum

  AUGUST 13, 14 and 15, 2018
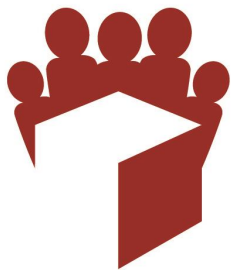
- Working groups are develop chapters for the publication

Join the conversation & add your expertise...

Wiki

Google Group

CD3DP.org

community standards
for **3D data** preservation