University of Oklahoma Libraries
DRAFT: Preserving Research Data at OU: From Text to Virtual Reality
Revised 3/10/17, Version 2.0

by Zack Lischer-Katz
Council on Library & Information Resources (CLIR)
Research Fellow in Data Curation

Table of Contents

Introduction

Campus-Wide Research Data Surveys at Other Universities

VR Research Data as Test Case

VR Research Data at OU: Where we are now

VR Research Data at OU: Where we are going

Problem Areas
- Suitable VR File Formats for Archiving
- Metadata for VR
- VR Repository Infrastructure
- Integration within Campus-wide Data Curation Strategy
- Preserving the OVAL Platform

Directions Forward

Research and Development for VR-related Data

OVAL Development and Preservation Timeline
- General VR Preservation Issues
  - Research and develop plan for preserving the VR platform
  - File formats for long-term preservation
- Metadata
- Repository
- Scholarly Tools for VR/3D

Tentative Presentation/Publication Deadlines:

References

Appendix A – Links to Existing Projects

Introduction

The storage and management of research data has become an increasingly important consideration across research universities. Data management plans are now required parts of federal grant applications, and data management is widely understood to be essential for promoting data sharing and reuse, protecting sensitive data (e.g. confidential human subjects data, health information) or intellectual property, supporting research reproducibility, and ensuring long-term data preservation and access. Ensuring the successful management and preservation of research data involves the establishment of data repositories, which include both social (policies, personnel, etc.) and technological (networks, servers, software, etc.) infrastructures necessary to ensure data integrity and documentation during data collection, processing, publication, archiving and reuse. Data repositories are are not only tasked with storing and maintaining the integrity of data, but must also document their activities and maintain the provenance of the data they store to maintain the context of data creation and ensure that datasets can be trusted by all stakeholders (researchers, funding agencies, the public, etc.).[1] Concerns over the reproducibility of research findings value and data reuse have lead to increased calls for data management plans and support infrastructure for a range of research disciplines.

As of December 2016, the University of Oklahoma is in the process of conducting a survey of the data management needs of its researchers. The findings of this survey will give empirical grounding for designing a comprehensive plan for data management and preservation at OU, which will have implications at the levels of policy, personnel and infrastructure. While it is important to apply technological solutions to address the growing data storage and preservation needs on campus, the institutional and social infrastructures of policy and expert personnel are also of critical importance.[2] Establishing the roles and responsibilities, data governance guidelines, and other policy-level frameworks will involve shaping university-level research policies and will

---

[1] In earlier eras of research, when researchers only shared data amongst their colleagues and research partners (which is often still the case in many research areas), the problem of "trust" in data was conferred through direct knowledge of the individuals involved. Contemporary researching that invovle data reuse and data sharing outside of the original research context now requires that each stage in the life of the data is documented to ensure trust in its authenticity and integrity (Ray, 2013). Reuse of data is unlikely to occur if researchers are not able to trust it.

[2] The library is well-positioned to play a central role in establishing a campus-wide data management infrastructure, as it is already sustaining a staff with the knowledge necessary to act as liaison between researchers and data management resources. Beginning to think about data curation as an extension of these services involves additional staff resources and training, but would begin to fill the existing knowledge gap that exists between researchers and data management and the library services required to support data management.

require allocating personnel to provide data curation services throughout the entire cycle of research.[3]

Campus-Wide Research Data Surveys at Other Universities

Until OU finishes conducting its own campus-wide research data survey, looking at the results of data surveys recently completed by other research universities may give some insight into what to expect in terms of the range of data types and the data storage needs of OU researchers. The following examples are drawn from surveys conducted on the research campuses of Cornell University, Oxford University, University of Iowa and Northwestern University (respondents were allowed to choose more than one data format, so percentages add up to greater than 100%):

- In 2011, Cornell University surveyed 84 of its NSF-funded Principal Investigators and found that they stored their data in a variety of forms:
    - 67.5% stored textual
    - 47.5% stored databases
    - 47.5% stored images
    - 43.75% stored computer code
    - 28.5% stored spreadsheets
    - 18.75% stored "other" (including:
        - 16.25% stored video
        - 10.13% stored audio
        - 10.13% reported "I don't know"

  Cornell also found most of the researchers' data were not "big data," but were medium-sized data, with roughly half of the respondents reporting that they needed less than 100 Gigabytes (GB) of storage for their data:
    - 2.38% reported data storage needs above 100 Terabytes (TB)
    - 10.71% reported data storage needs between 1TB and 100TB
    - 16.67% reported data storage needs between 100GB and 1TB
    - 30.95% reported data storage needs between 1GB and 100GB
    - 22.62% reported data storage of "no more than" 1GB

---

[3] As suggested in the Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010):

> The strongest incentives to preserve will be ineffective without explicit agreement on the roles and responsibilities of all the actors—those who create the information, those who own it, those who preserve it, and those who make it available for use. Every organization that creates and uses data should implement policies and procedures for preservation, including: selection of materials with long-term value; preparation of data for archiving; and protocols to ensure a smooth and secure transfer of digital assets across organizational boundaries and between institutions. (p. 2)

    - 16.67% did not respond to the question.

- In 2012, Oxford University conducted a survey (with 314 responses) that found:
    - “61% of survey respondents worked with textual data, 65% with numerical data, and 62% with statistical data. Perhaps surprisingly, more specialized data types were also quite widely collected. 43% of respondents had image data, 15% had audio data, and 14% worked with geospatial data. 76% of respondents used spreadsheets to store at least some of their data, whilst 32% used relational databases” (Wilson, January 2013).

- The University of Iowa also conducted a survey in 2012 (with 784 respondents) and found:
    - “Most participants reported use of tabular/relational data (404 responses, [51.5%]) or textual data (391 responses, [49.9%]). […] 42 [5.4%] respondents reported collecting, generating, or using other types of data than the options listed. Some of these free-text responses, such as “physical measurements”, “interviews and field notes,” and “time series data,” represented refinements of the options given, while others such as “instrument or telescope data” and academic and medical records were types that did not fall cleanly into any of the options offered” (Averkamp and Gu, 2012, p. 7).

- In 2015 Northwestern University conducted a survey (with 831 responses and a response rate of 6.4% of its researchers, including faculty and graduate students) with the following results:
    - “The survey shows that research data comes in many different forms. Most common data types include spreadsheets (68%), structured data (e.g. csv, xml) (58%), text (74%), and images (52%). [...] [10%] of respondents selected “Other” data types. These included crystallography data, mathematics, a custom format used by the lab, historical archives, various types of experimental measurements (EEGs, NMR spectroscopy, seismic data, medical image data, and genetic sequencing data files)” (Buys and Shaw, 2015, pp. 9-10).

  The results of the data survey conducted at Northwestern in 2015 also points to the difficulty in scoping the size of a data repository for all types of data:
    - “the different types and sizes of data indicated by the respondents show that a universal data storage/preservation solution will be difficult. Estimation of university-wide storage solutions is also difficult as 31% of respondents did not know the size of their data storage needs.

> Additionally, many respondents indicated that they planned to keep their data indefinitely. Any institutional storage solution will need to accommodate many data types and uncertain storage capacity needs over long periods of time" (Buys and Shaw, 2015, pp. 15-16).

These four examples taken from other research universities suggest that although most researchers store their data within simple formats (text, spreadsheets, structured data, and images) and most data storage requirements are below 100GB of storage space, campus-wide research data can contain significant quantities of complex formats, such as video, audio, 3D objects and other data types that require careful attention for data curation and preservation.

Assuming that OU faces a similar data landscape as other research universities, these surveys suggest that OU researchers across campus likely work with a variety of data types, with the vast majority working with relatively small data sets, the vast majority of which are in the form of text files or structured numerical data. While these types of files are comparatively simple and relatively small in terms of the storage space they require, the surveys also suggest that a significant number of researchers work with complex data types, and that the metadata required to document them may be very project-specific. There are also some suggestions by library researchers that libraries need to support both big and 'little' data projects[4]. In terms of big data there are a number of data formats particular to OU that will require special consideration due to their size or complexity, such as meteorological data, which are very large in scale. Additionally, there are emerging formats, such as 3D models, 360 video and virtual reality platforms being hosted at OU Libraries as a platform for research and pedagogy that require special attention because they are complex and the metadata standards and data management practices for them are not fully established. Properly managing

[4] Joyce Ray (2013) suggests that libraries may be best positioned to address the needs of smaller research projects since they lack the large-scale infrastructure needed for big data management and preservation, and rather than "big data" projects, many researchers use a wide range of types of "small data" that may appear in non-standardized ways, making management more difficult:

> Data that result from smaller projects often are more difficult to manage than big data because they are highly heterogeneous, require more individual attention per byte, and tend to be less well documented. Academic libraries generally lack the capacity to manage the large volumes associated with big data, but they may be well equipped to assist with managing smaller data projects. For example, they may recommend sustainable file formats and file organization, advise on intellectual property issues for data reuse, assist with determining appropriate metadata and data citation practices, and provide repository services for managing current research data as well as for archiving data after project completion. (p. 2)

and curating data related these emergent technologies will likely have an impact on whether these new data visualization tools become adopted into faculty research.

This latter category of data, virtual reality-related data, is an emergent area of data production that may offer an important test site for many of the issues that face data management and preservation at OU, more generally.

VR Research Data as Test Case

For emerging technologies, exemplified by virtual reality, standards and best practices are not well established in the field. This suggests that OU Libraries has an opportunity to take the lead in developing standards for these types of research materials. In addition, the complexity of the data objects, in terms of variety of data types, relations between files, and dependencies on a variety of hardware and software platforms, makes it a particularly difficult format to manage and archive, offering a challenging test case for any data repository design. Thus, it is argued that using VR-related data as a test case will be helpful for identifying key barriers to implementing a data repository that will able to sustain the range of other types of data on campus. In the remainder of this report, the current state of virtual reality-related data on campus will be described and future directions outlined.

VR Research Data at OU: Where we are now

The Oklahoma Virtual Research Laboratory (OVAL) is growing rapidly as a pedagogical and scholarly tool. At present, OVAL hosts a variety of 3D model file types within an immersive virtual reality (VR) environment, and the following disciplines and organizations are engaged in developing OVAL for various applications: Law, Interior Design and Architecture, Biology, Chemistry/Biochemistry, Informatics/Computing, Medical Imaging, Development, Sam Noble Museum of Natural History, Art and Art History, English, Earth & Energy Research, Physics, the Oklahoma Mesonet, Geology, Journalism, Anthropology, Cognitive Studies, and Engineering. This wide range of current and potential users and uses of OVAL demonstrates the strong interest across campus in this networked VR platform that can support a range of research and pedagogical applications and is hosted within a library context. The location of VR at the center of academic life, the library, gives OU a unique opportunity to provide VR and VR-related technologies as innovative new tools for scholarship across the disciplines. Interest in developing VR technology for research and pedagogy also exists outside of OU, with interest being expressed by other academic institutions to adopt OVAL, including the University of Arizona and North Carolina State University. This suggests that forming partnerships and encouraging adoption of the VR platform beyond the OU campus would be an important strategy for supporting OVAL, and expanding its scholarly impact in the long-term.

VR Research Data at OU: Where we are going

OVAL has already proven itself to be a useful tool for a number of academic departments, and - more broadly - as a device that consistently piques interest in VR technologies for many more. Some departments, such as Architecture, English, Anthropology and Chem/Bio-Chem are currently implementing OVAL into course assignments, and research collaborations between faculty and OVAL team members have already lead to conference presentations and co-authored paper publications. At the same, it is becoming apparent that OVAL needs to expand its functionality and add several key research tools and infrastructural components to produce a more robust academic platform. Part of this may be accomplished through the implementation of additional resources for OVAL's development, as well as through research projects carried out by the OVAL team to evaluate the impact of OVAL on scholarship and teaching, and identify new scholarly area open to using OVAL. Taking stock of the overall flow of data through the OVAL platform and the various people, tasks and workflows related at each stage would also likely help to keep track of the various components of OVAL and determine where metadata and other methods of documentation may be needed to ensure the long-term preservation of OVAL and the research products that produced through its use. Developing an overall data management plan would help structure the documentation at each stage in OVAL. Files and associated metadata need to be documented and stored systematically so that they may be eventually integrated with any future campus-wide data repository.

In addition, a range of new tools for OVAL are being assessed by the OVAL development team, based on feedback from current OVAL users. The OVAL team has either discussed or received feature expansion requests related to: 3D visualization of numerical data for data exploration (interest has been expressed by Economics and finance researchers, and there are clear applications for meteorology and geology, as well), adding a robust toolkit that can enable research functions such as annotation (text, visual, audio), precise measurement (Anthropology professors are already interested in using this tool in class assignments), object slicing, animation, version control, video documentation and creating scenes of multiple 3D models, and enabling the export of various data types for publication (3D-PDF, video files, CSVs of measurement data, etc.).

Going hand in hand with the development of new tools is the development of a suitable VR data repository that can support the metadata and master file formats necessary for long-term preservation, and that can manage the derivative files that students and faculty search for (or upload) in order to access in OVAL. This repository will also likely need to support the new mobile OVAL platform currently under development, and will require the recording of metadata throughout the entire lifecycle of the digital models, including provenance metadata on the creation of the 3D objects, metadata linking together associated annotations and measurements, and information about the version of the VR platform the researcher used to view and work with the objects. For instance, our contacts at the Law School Library are very concerned about issues of chain of custody, authenticity and security around VR content used as evidence for legal proceedings, and similarly, provenance metadata, as it records the precision of scans and the methods used to produce and process the scans of artifacts and locations is important for archaeology, anthropology and other disciplines that work with human artifacts and locations.

At the same time, we are working to develop instruments (e.g., surveys) for gathering data from students who are using OVAL for class assignments in order to evaluate OVAL’s impact on student learning and visual literacy, as well as gathering feedback from faculty partners about toolset functionality. Interest from professors in Chem/Biochem, as well as Anthropology have shown support for carrying out the research necessary for evaluating OVAL’s pedagogical impact. We would also like to develop a protocol for interviewing and/or surveying faculty members to gain new knowledge about how they are using (or might hope to use) OVAL, in order to increase understanding of the range of domain-specific research practices and how they may be supported in a VR environment, and to evaluate the usability and functionality of OVAL as a research tool.

Problem Areas

In light of the vision for the future of OVAL outlined above, the following problem areas have been identified as key research areas that will require careful consideration: Choosing a sustainable file format for the long-term preservation and use of 3D models; developing a comprehensive metadata schema; building a VR repository infrastructure; taking into account how the OVAL repository will fit in with the greater campus-wide data curation initiatives; and developing a plan for the long-term preservation of the OVAL platform, including documenting any modifications made to its software and hardware configuration.

Suitable VR File Formats for Archiving

The selection of sustainable file formats is important for ensuring data are not lost in the future due to format obsolescence. The current 3D file format used in OVAL is the .obj format. It is widely adopted and is an open and easy to understand format. Data within .obj files are stored as sets of positions and vectors, which can be opened and examined in a text editor, if need be. A significant limitation of the .obj format is that it only contains the information for the polygon mesh that defines the structure of the 3D object. .obj files require external files for storing texture and color information, through the use of .mtl and .jpg (or .png) files. This requires the management and preservation of three separate files to preserve a basic 3D model, with structure, color and texture (other attributes, such as annotations, animation, etc., may require additional files). .obj is also not an ideal file format because there is no standard way of embedding metadata in the file. An additional file containing the metadata must be stored alongside the three other files. While the problem of storing four total files per digital object and forming relations between those objects in a digital repository is certainly solvable, a more robust file format that can store texture, image and metadata may be worth seeking out. The inability to embed metadata within .obj files poses problems for precisely measuring objects in 3D space, since the scale information is cannot be stored within the .obj files.

Current work we are pursuing on this problem includes investigating the usefulness of alternative file formats, such as the open XML-based formats, .x3d (superseding VRML) and the Collada format, as well as the .ply file format, and the proprietary .fbx file format. Key concerns around the evaluation of these formats center on the ability to embed technical metadata, including precise scale and dimensional information, and provenance metadata. Also of concern is the flexibility of the format to support future extensions of the OVAL system, such as animations, scene construction and physical properties (gravity, mass, velocity, etc.). The long-term sustainability factors of each 3D file format are also being assessed, using criteria established by the Library of Congress (http://www.digitalpreservation.gov/formats/sustain/sustain.shtml).

Metadata for VR

Developing a comprehensive metadata schema is essential for for being about to describe and retrieve 3D objects, to document provenance and technical aspects of the file creation process (from scanning to processing to output), and track intellectual property rights and usage restrictions. In addition, some metadata elements (such as scale and measurement information) are integral to supporting extended OVAL functionality, such as precision measurements.

A review of the existing literature shows that there is no standard metadata schema for 3D models. The following schemas considered together may be useful for supporting some of the metadata needs of an OVAL repository: CRM-dig (provenance metadata schema for documenting the production of digitized 2D and 3D objects); VRA Core or

CARARE 2.0 schema (descriptive metadata schemas for visual materials; CARARE 2.0 was developed with 3D materials in mind, so it might be preferable); and METS (structural metadata describing relationships between different files and resources) (see Appendix A – Links to Existing Projects for links to more information).

Further investigation and consultation with metadata experts will be necessary, and in order to embed metadata into the file format, a suitable file format will need to be selected, as discussed earlier. The MPEG-V standard (released in 2011), which defines an exchange schema for information between virtual world environments, may also be worth investigating as a potential schema for documenting the OVAL platform itself, with implications for its long-term sustainability.

VR Repository Infrastructure

A robust VR repository infrastructure would support both the management and preservation of research data as well as the access and use of that data by OVAL users. We are currently developing a data catalog and cloud-based storage solution for managing 3D files, storing metadata and producing archival submission packages. This is a medium-term data repository that will act as a platform upon which we can build workflows and metadata schemas to manage and document 3D creation for scholarship, scholarly and pedagogical uses in the OVAL platform, and manage the collections for digital preservation, including monitoring file integrity and tracking risks of file format obsolescence over time.

There has also been some discussion with other institutions about developing a shared a repository of 3D objects. Interest has been shown by librarians at the University of Arizona, North Carolina State University and Indiana University. Nurturing these partnerships could produce a useful pooling of resources that could help to develop one standard repository structure and metadata schema. It will have to be sorted out if these partnerships will yield a common repository infrastructure or if the end goal would be a search portal that would enable federated search across each institution’s local repositories. We also been in conversation with members of the University of Arkansas and their Center for Advanced Spatial Technology (CAST) about working together to develop repository and metadata guidelines. CAST has many years of experience carrying out 3D scanning and other advanced spatial data collection projects, and it has been an excellent resource for answering questions about the 3D model creation process. CAST has also agreed to let OU Libraries host some of its content on the OVAL platform, which offer an excellent test case for considering how 3D files can be effectively transferred between institutions.

Intellectual property and privacy are issues that have been brought up about OVAL. One solution could be to implement an authentication layer to restrict access and enable OU users to login using their OU 4+4 credentials. This would help ensure the protection of sensitive or restricted 3D models, and by maintaining a database of user profiles, users could login and have a workspace that would only contain the models that they have uploaded (or a faculty member has given them access to). This will likely become a particularly critical component of OVAL when a mobile platform is developed (currently in works), and addresses questions about logging in from off campus, and concerns over limiting storage and bandwidth for mobile viewing of 3D models. Faculty may also feel more comfortable uploading their authored 3D models if they can be assured that only they (and their students) will have access to those models. Similarly, our partners at the OU Law School Library have expressed interest in the ability to restrict access to models of crime scenes. Under this authentication / user profile scheme, faculty could restrict access to their models to only the students in their class, or share with other researchers. Developing this component will likely need coordination with the current work being done to develop an authentication layer for special library resources.

Integration within Campus-wide Data Curation Strategy

Given the OU Library’s initiative to support data curation for a wide-range of types and sources across the OU campus (including data from partner institutions), any data repository for 3D objects and other VR-related data should be designed in concert with ongoing efforts to develop a comprehensive data repository. A medium-term, flexible data repository is needed that can then be integrated when a campus-wide data repository is fully operational at a later date. The medium-term repository will need to be able to store and manage metadata and files in a way that will enable the collection to be easily transferred into a comprehensive data repository in the future.

Preserving the OVAL Platform

Considering the preservation of the OVAL platform, since it shares some properties with an interactive virtual world, and shares many of the same concerns related to the preservation of video games and complex computer software/hardware configurations. The following nine problems associated with the long-term preservation of video games were identified in the University of Illinois’s *Preserving Virtual Worlds Project Final Report* (McDonough, et al., 2010, pp. 14-15):

1. Hardware obsolescence: “The original console or computing platform used to run the game may cease to be supported or even available in the aftermarket.”

2. Software obsolescence: “The original software needed to run the game—operating system, drivers, frameworks—may lose support, cease development, or become incapable of running on future hardware/software configurations.”
3. Scarcity: “Some video games are produced in limited quantities and are subject to the dangers of media decay.”
4. Third party dependencies: “Currently most emulators are developed by the game community and are of questionable legality.”
5. Complex, proprietary code – And lack of documentation: “Digital games are generally released as compiled binaries with no documentation of the compiling process, or even the programming languages used.”
6. Authenticity: ”Proving that a digital object is what it claims to be, free from tampering or corruption.”
7. Intellectual Property Rights: “The game development industry is highly creative and competitive, leading developers to be conservative with their intellectual property.”
8. Significant properties: “What are the significant properties of a game that must be maintained with each transformation/preservation action? ... How faithful must we stay to the original code? Significant properties are essential to define, as they play a major role in determining authenticity.”
9. Context: “Although not an immediate threat to the preservation of games, building contextuality is important to creating understanding for future users.”

From this list of problems identified in the *Preserving Virtual Worlds Project Final Report*, the ones seemingly most relevant to OVAL are:

1. Hardware obsolescence: What is the impact on scholarship when Oculus or Leap Motion change their hardware in ways that change the ways in which OVAL behaves?
2. Software obsolescence: What happens when Unity changes its software in ways that breaks existing functionality of OVAL?
3. Complex, proprietary code and lack of documentation: Will Unity give us the source code for earlier, obsolete versions of its software?
4. Authenticity: How do we ensure that 3D models in OVAL are being represented accurately and have not been tampered with? This is particularly important when scientific measurements are being conducted on a model, or if the model represents evidence in a courtroom.
5. Intellectual property rights: In what ways do software licenses limit what we can do to archive commercial software? This is also important to think about when we

are working with 3D models in OVAL that are protected by intellectual property law. How do we ensure that we aren’t violating any of the licenses associated with the 3D models or that we are not inadvertently making publically accessible faculty member’s proprietary research data?

6. Significant properties: What properties of the OVAL platform are significant to users? How do we ensure that these aspects are not changed or altered when updating software, emulating an earlier operating environment or migrating to a new platform?
7. Context: How do we document the research or teaching environment in which the model was built, viewed and/or modified? What will be important for future users, library administrators, and historians to know about the earlier versions of the OVAL platform?

Some possible strategies for addressing these barriers facing the long-term sustainability of OVAL and documenting its significant properties include:

- Maintaining a repository of the earlier versions of the OVAL software and hardware components, and documenting in each digital object’s technical metadata the platform configuration that it was viewed and manipulated on. For scholars publishing results based on OVAL-supported 3D visualizations, this information would also be important for the researcher to include in their final publication or other research output. Acquiring copies of the binaries or the source code for earlier versions of Unity or any hardware drivers would also help ensure that files used in earlier system configurations of the OVAL platform could still be accessed into the future.
- Outputting video files (either 2D or 3D) that document user manipulations of the 3D models in the OVAL platform. This has the advantage of both preserving the look and feel of the OVAL platform, but also documenting the actual interactions of a researcher with the system. A 3D video output would also allow viewers to have some degree of 3D motion in the playback of the recorded OVAL session.
- Containerization of the OVAL platform and implementation as cloud-based software could be one solution for the long-term preservation. As with the mobile OVAL platform currently being developed, there would also be questions of bandwidth and latency if OVAL were implemented as a cloud-based solution. If any critical components of Unity, Oculus, Leap Motion, 3D mouse, or any other components of the OVAL platform begin to become unsupported by its manufacturer, then containerization and emulation becomes to seem like a more attractive option. This does not entirely circumvent the hardware preservation

issue, which can be addressed by developing the institutional knowledge to repair them, and acquiring backup hardware.
- Beyond the usual challenges of software preservation, preserving OVAL is complicated by the fact that it is still under development. Because OVAL’s functionality and its user base are all still evolving, there exists a significant degree of uncertainty about what properties of OVAL will be seen as “significant,” and will need to be preserved. Thus, developing a preservation plan for the OVAL platform will likely be iterative, changing with the evolution of the OVAL platform and its users. Any preservation plan, like a data management plan, may best be described as a living document that can change as new data types, uses and stakeholders emerge over time. .

Directions Forward

In summary, the following directions forward in making OVAL more usable and preservable are recommended:

- We should continue experimenting with alternative file formats in OVAL, evaluating their capacities for embedding metadata and testing their ability to extend functionality in OVAL, and assessing sustainability factors of each potential format.
- We should continue consulting with metadata experts inside and outside OU libraries, and evaluate suitable metadata schemas for fully documenting the various aspects of 3D model production, use and preservation. This will be supported by the development of a 3D model production workflow and the identification of key points in the production process that may require metadata collection. A metadata schema for OVAL should be able to document provenance, description, structure, rights, and technical metadata, and be able to provide satisfactory documentation for all types of users and uses.
- Begin to develop a data management plan for 3D data used in OVAL, and begin to establish selection criteria for what should be archived for long-term preservation. Currently, most of the 3D models being used in OVAL are downloaded by students or faculty from existing 3D repositories, but this is changing. Faculty and students are beginning to author their own models, and the OVAL team is beginning to do its own 3D scanning using laser scanning and photogrammetry. It will be necessary to work with domain experts to develop appraisal policies for identifying 3D models in need of short-, medium- and long-term data management. Data linked to by publications will require special attention. Identifying the various points in the research process in which 3D data may be produced, processed and analyzed will require additional research. For instance, a researcher may load a numerical data set or 3D model into VR,

manipulate the visualization in some way, and then publish the results of that analysis in a paper. We will need to investigate which data are necessary to archive at different points in the research process in order to support “rigorous” scholarship (defined in terms of discipline-specific conventions and expectations).

- The OVAL team should consult with other organizations that have more experience carrying out 3D scanning projects and have already developed 3D repositories and workflows. Considerable work has been done over the last decade on artifact and architecture scanning, repository structures and metadata schemas for 3D objects, including work done at the University of Arkansas (Fred Limp, Center for Advanced Spatial Technologies); Arizona State University (the Digital Archaeological Record); Europeana (3D-ICONS project); Archaeology Data Service (University of York); 3D Coform Initiative (University of Brighton, with funding from EC's Seventh Framework Programme); the CIDOC-CRM framework (International Council of Museums, with European Commission funding); and Cornell University’s Guidelines for Preserving Digital Art Objects (see Appendix A – Links to Existing Projects for links to these projects). In December 2016, the OVAL team visited the Center for Advanced Spatial Technology - University of Arkansas (http://cast.uark.edu/) where we were provided with two days of 3D scanning demos and tours of the facility, and we are now nurturing a partnership that could lead to collaborative work on the preservation issues associated with 3D scanning and will invovle the sharing of 3D content. It would be helpful to look at these other projects and engage with their key personnel for guidance on the development of 3D scanning workflows, repository structures and metadata through:
    - Examining their published reports.
    - Consulting directly with experts.
    - Planning a symposium to bring experts to OU to share knowledge from earlier 3D repository initiatives (in the planning stages with other five other CLIR fellows).
    - Conducting site visits at other institutions.
- Continue recruiting other academic libraries to adopt OVAL and become part of our development community. Active recruitment could be supported by a wiki, discussion board, or other social media application that would provide a forum for the growing OVAL development community.
- Identify grant funding opportunities that could be used to document and package OVAL for easy adoption at other universities, develop new tools for OVAL, and develop specifications for a 3D repository and expand on our existing 3D

production abilities through training and purchase of more advanced scanning hardware and software.

- Continue to research the risks and advantages associated with the creation of a shared repository among other university libraries, and continue conversations with interested institutions in order to encourage outside adoption of the OVAL platform.
- Consult with experts in the field of video game / virtual world preservation, and software preservation to better understand the challenges of preserving the OVAL platform over the long-term. Consult with experts on the possibility of establishing standards that might benefit the wider community.
- Contact the commercial hardware and software corporations that developed the components that OVAL is built on, including Oculus, LeapMotion, and Unity. These companies may be able to offer advice or resources on developing and preserving these components as part of a larger OVAL preservation plan.

Research and Development for VR-related Data

The research and development needed to address these concerns for the management of VR/3D research data has already begun. Individual meetings have started happening between the OU team and groups at the University of Arkansas, Center for Advanced Spatial Technology, North Carolina State University, Indiana University, the University of Arizona, and with other potential external partner institutions. These meetings have started to address some the of the major technical questions around assessing sustainable 3D file formats, metadata schemas, 3D asset repositories, content sharing, and sharing OVAL and expanding its tools.

Moving ahead, the following section outlines a timeline containing key steps in the process of research and development around OVAL in terms of extending its capabilities for research and pedagogy, followed by a list of potential venues and deadlines for presenting on the outcomes of this research.

OVAL Development and Preservation Timeline

General VR Preservation Issues

<u>Research and develop plan for preserving the VR platform</u>

Research existing literature (Sept. 2016 - Dec. 2016)
Consult with experts in the field (Oct. 2016 - Oct. 2017)
Develop preservation plan (Dec. 2016 - Oct. 2017)

<u>File formats for long-term preservation</u>

Research existing literature (Sept. 2016 - Dec. 2016)
Consult with experts in the field (Oct. 2016 - Oct. 2017)

Analyze file format sustainability (Nov. 2016 - Oct. 2017)

Metadata

Develop Core Technical Metadata Schema (Nov. 2016 - March 2017)
Develop Workflow for Technical Metadata in 3D Scanning (Nov. 2016 - March 2017)
Develop Core Descriptive Metadata Schema (Dec. 2016 - April 2017)
Develop Modules for Domain Specific Metadata (Jan. 2017 - June 2017)

Repository

Review Literature on VR/3D Repositories (Sept. 2016 - Dec. 2016)
Research Potential Partnerships for VR/3D Repositories (Oct. 2016 - April 2017)
Plan Medium-term Data Repository for VR/3D (Oct. 2016 - Feb. 2016)
Construction of Data Repository (Feb. 2017 - June 2017)

Scholarly Tools for VR/3D

2D/3D Recording (Dec. 2016 - March 2017)
Annotation (Jan. 2017 - Feb. 2017)
Measuring Tools (Feb. 2017- April 2017)
Animations (April 2017 - Sept. 2017)

Tentative Presentation/Publication Deadlines:

1. *Paper Working Title:* “Preserving Virtual Reality Data in a Research Library” – Define value of VR in research library context and outline problems and the steps we are taking to solve them.

a. Submit Presentation Proposal for CNI (April 3-4, 2017)(Deadline TBA)

b. Revise based on feedback and submit to Library Trends (or equivalent) (submit by end of April 2017)

2. *Paper Working Title:* “Metadata Schema for Preserving Virtual Reality Data in a Research Library Context” – Report on results of research on developing a metadata schema for preserving virtual reality data.

a. Submit to Archival Science (Spring 2017)

3. *Paper Working Title:* Visual Epistemologies of Virtual Reality: 3D Models as Research Data

a. Submit to: Center for Material Culture Studies Symposium, “Imagined Forms: Modeling and Material Culture” University of Delaware (Nov. 17-18 2017). (Deadline: Feb. 15, 2017)

b. Revise and submit to Journal of Visual Culture (Fall 2017)

4. *Other conferences to submit to (deadlines TBA):*

a. Association of Information Science and Technology (Fall 2017) (deadline ~April. 2017)

b. DLF Conference (Fall 2017) (deadline: ~May 2017)

c. Association for Library and Information Science Education (Winter 2018) (deadline ~July 2017)

d. iConference (Spring 2018) (deadline ~Sept. 2017)

References

Averkamp, S., & Gu, X. (2012). Report on the University Libraries’ data management needs survey. University of Iowa Libraries Staff Publications. Retrieved from http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1245&context=lib_pubs

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2010). Final report: Sustainable economics for a digital planet: Ensuring long-term access to digital information. Retrieved from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Buys, C. M., & Shaw, P. L. (2015). Data management practices across an institution: Survey and report. *Journal of Librarianship and Scholarly Communication, 3*(2), eP1225. http://dx.doi.org/10.7710/2162-3309.1225

McDonough, J.P., Oldendorf, R., Kirschenbaum, M., Kraus, K., Reside, D., Donahue, R., … Rojo, S. (2010). Preserving Virtual Worlds: Final report. Retrieved from https://www.ideals.illinois.edu/handle/2142/17097

Steinhart, G., Chen, E., Arguillas, F., Dietrich, D., & Kramer, S. (2011). Prepared to plan? A snapshot of researcher readiness to address data management planning requirements. *Journal of eScience Librarianship, 1*(2). Retrieved from http://escholarship.umassmed.edu/cgi/viewcontent.cgi?article=1008&context=jeslib

Ray, J. (2013). *Research data management: Practical strategies for information professionals.* West Lafayette, ID: Purdue University Press.

Wilson, S. (2013 January). University of Oxford research data management survey 2012: The results. Data Management Rollout at Oxford (DaMaRO) project blog. Retrieved from http://blogs.it.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-the-results/

Appendix A – Links to Existing Projects

3D Coform Initiative, University of Brighton, https://www.brighton.ac.uk/research/our research/life-health-and-physical-sciences/research-groups/computing/3d-coform.aspx

3D-ICONS project, Europeana, http://pro.europeana.eu/project/3d-icons

Archaeology Data Service, University of York, http://archaeologydataservice.ac.uk/

Center for Advanced Spatial Technologies, University of Arkansas, http://cast.uark.edu/

CIDOC-CRM Framework, International Council of Museums, http://cidoc-crm.org/

The Digital Archaeological Record, Arizona State University, http://www.tdar.org/about/

Guidelines for Preserving Digital Art Objects, Cornell University, https://ecommons.cornell.edu/handle/1813/41368

Preserving Virtual Worlds Project, University of Illinois, https://www.ideals.illinois.edu/handle/2142/17097

Metadata Schemas

CARARE 2.0, http://3dicons-project.eu/eng/Resources/Documentation/CARARE-2.0-schema

CRM-dig, http://www.ics.forth.gr/isl/index_main.php?l=e&c=656

METS, http://www.loc.gov/standards/mets/

MPEG-V, http://mpeg.chiariglione.org/standards/mpeg-v

VRA Core, http://loc.gov/standards/vracore/

3D File Formats

Collada, https://www.khronos.org/collada/

FBX, http://www.autodesk.com/products/fbx/overview

MTL, http://paulbourke.net/dataformats/mtl/

OBJ, http://www.martinreddy.net/gfx/3d/OBJ.spec

PLY, http://paulbourke.net/dataformats/ply/

STL, http://www.fabbers.com/tech/STL_Format

X3D, http://www.web3d.org/x3d/what-x3d

Potential Partners

Indiana University, https://libraries.indiana.edu/virtual-reality

North Carolina State University, https://www.lib.ncsu.edu/do/virtual-reality

University of Arizona, http://new.library.arizona.edu/visit/spaces/ispace

University of Arkansas, Center for Advanced Spatial Technology, http://cast.uark.edu/