

Jamie Wittenberg

New dimensions in data management

Understanding sharing and reuse practices for 3-D data

As 3-D digitization becomes more accessible and research institutions expand support for 3-D modeling, researchers are increasingly leveraging 3-D models and methods. For instance, a paleontologist might use a micro CT scanning process to capture images of the inside of a specimen that would otherwise be destroyed by such an analysis. An archaeologist might use photogrammetry to construct digital representations of artifacts that can then be examined in a way that would be difficult or impossible in a museum setting. The emergence of 3-D modeling as a research practice presents several challenges for libraries working to support and facilitate the dissemination and reuse of 3-D data packages. At present, there is significant work to be done in the community to create a culture and infrastructure that facilitates sharing 3-D research.

Understanding data sharing and reuse among researchers is critical to the success of collection, dissemination, and preservation efforts among memory institutions. Existing literature on data sharing, reuse, trust, quality, and review can inform approaches to evaluating how researchers might share or reuse 3-D data. However, 3-D data have characteristics that make them unique—rapidly changing technology, intersections with lucrative commercial sectors like virtual reality gaming, and the expectation that a model will render—when

shared. This project offers a unique and necessary contribution to the literature in analyzing creation, reuse, and publishing of 3-D through interviews with expert researchers. This provides substantial value to libraries, archives, and museums that work with 3-D by enabling memory institutions to design digital collection and repository systems that meet patron needs and foster innovation.

Methods

Participants were recruited to the study in one of three ways: 1) They authored a relevant and impactful article on 3-D methods; 2) They participated in one or both IMLS-funded 3-D forums; and 3) They were referred by another interview participant. Twelve semi-structured interviews were conducted ranging from 45-to-75 minutes. All interviews were recorded and transcribed prior to analysis.

Of the 12 expert researchers interviewed, four self-identified as archeologists, four self-identified as paleontologists, and four self-identified as collection managers. These designations are somewhat fluid, as three out of the four collection managers participating in the study have

Jamie Wittenberg, formerly of Indiana University-Bloomington, is now assistant dean for digital strategy and IT at University of Colorado-Boulder Libraries, email: jamie.wittenberg@colorado.edu.

© 2020 Jamie Wittenberg

graduate degrees in a field related to archeology or biology. The interview questions were designed to elicit responses that would provide a basis for answering the following research questions:

- What are current practices and standards for sharing and reusing 3-D data?
- What are researchers' requirements of 3-D data packages that are (or could be) incorporated into their own research?
- How important is sharing a 3-D model to scholarly communication, and how satisfied are researchers with the infrastructure that supports sharing?

Findings

Discussions related to sharing 3-D data packages are largely discipline-specific and have given rise to key standards development work and repositories. When asked, "What happens to your 3-D objects when your study is complete?" 50% of respondents stated that at least some of their data are deposited into an institutional or disciplinary repository. While 50% of respondents indicated that they both create their own models and use models created by peers, the remainder only used models in their research that they had digitized and/or constructed themselves. When respondents did use models created by other researchers, they were often created by another member of the research team or a researcher in the same scanning facility. Twenty-five percent of respondents specifically mentioned downloading datasets from repositories.

One paleontologist stated that "despite the high cost of digitization of biological specimens, it is almost always more worthwhile to do it yourself or by a trained morphometrician rather than utilize somebody else's data." The respondent cited concerns about documentation and precision. There was one notable exception to this—one participant shared that they reached out to an organization that had digitized an artifact in a remote location. This object would have been cost-prohibitive for the

respondent to capture. After some persistence, the organization was convinced to share the model. The data were shared successfully in Dropbox in multiple formats -- something the respondent highlighted as important in this kind of sharing because one of the formats was unusable to their research group. The respondent was able to incorporate the shared data into their analysis and treat it in the same way as the models digitized in-house.

Participants were asked the question: "If you were to access objects created by others and incorporate them into your research, what information would you need in order to make those objects useful?" Even participants who reported that they have not or would not reuse 3-D data generated for another purpose were asked to provide a hypothetical response. This question evoked perhaps the greatest diversity in perspectives.

One archeology researcher reported that for their analysis, it would be helpful to know what kind of instrument was used, but aside from that they would have no reuse requirements because it would likely not make a statistical difference. A researcher in paleontology expressed that they would only need to know the dimensions of the voxels and contextual information about the specimen, such as sex and where it was collected. Their perspective was that they would only need access to the raw data if something about the slices "looked wrong."

Other respondents stated that they would need extremely precise information about the object and digitization process, including GIS coordinates, instrument configuration, alignment, paradata, provenance data, technical specifications, the "rawest form of the data," scale information, StereoLithography and Wavefront 3-D Object files, readme file, quantification of data cleaning, published accuracy of scanner, type of scanner, details of post processing, name and affiliation of individual who scanned the object, purpose

of digitization, and metadata in the appropriate schema.

When asked about the role of 3-D in their research, many respondents reported that 3-D digitization is an integral and central part of their analysis. One researcher stated, “I would say the 3-D objects are fundamental to my research and are the entire point of research.”

Both paleontologists and archaeologists frequently cited the importance of 3-D in examining aspects of specimens and artefacts that could not otherwise be analyzed. For example, paleontologists reported that volumetric data are critically important and can be achieved through CT scans. Archeologists highlighted reconstructions of ancient civilizations, monuments, and objects. Curators and collection managers emphasized facilitating access to physically sensitive materials. Respondents noted that 3-D models allow access to objects that would otherwise be destroyed as part of the analysis. Many respondents expressed that though they consider sharing models very important in communicating their research, they were not satisfied with their options for sharing 3-D data packages.

Overall, respondents ranked the importance of sharing 3-D data an average of 4.42 and infrastructure for storing and sharing an average of 2 on a scale of 1-5, with 1 being the least satisfied and 5 being the most satisfied.

There were several universal trends that emerged around the questions of challenges associated with 3-D methods, irrespective of respondents’ discipline. Fifty percent of experts interviewed identified sustainability, reproducibility, and documentation as a challenge.

One respondent, in addressing this challenge, noted that in the process of constructing a 3-D model, “at every point [there is] decision making about choosing what to show [or] what not to show” and asserted that researchers using 3-D methods need to improve documentation

of these decisions in order to facilitate reproducibility. One respondent shared that “everyone [in my field] is worried about archiving for reuse [and the] loss of irreplaceable field data.” Another participant said, “the ability to take those raw [3-D] data from even sort-of long tail science projects is abysmally non-reproducible.”

One of the most commonly referenced challenges is the cost of 3-D research methods. Ninety-two percent of researchers interviewed said that the cost of high-quality digitization equipment posed a challenge. Several researchers interviewed noted that photogrammetry is becoming increasingly popular because it is significantly less costly than other methods of 3-D digitization. Even among respondents with access to well-funded and established digitization facilities, cost was cited as a challenge. Another frequently referenced challenge is difficulty scanning certain kinds of objects, including reflective objects, extremely dense objects, and living organisms that shift position.

Overall, respondents expressed concern about data ownership and copyright. Respondents frequently expressed confusion about rights issues and identified copyright considerations of 3-D as an area that needs attention from intellectual property experts. Though copyright and rights issues were the primary access restrictions identified by respondents, other important barriers to sharing were also discussed. Respondents across disciplines highlighted ethical and legal considerations when constructing models based on indigenous artifacts, monuments, or other objects. One researcher stated that work needed to be done “finding a happy medium between being respectful but still allowing scientific inquiry.” Respondents expressed concern about 3-D models of objects in museums being illegally disseminated or sold, even when licensed only for noncommercial use. Respondents noted that this kind of activity makes it more difficult for researchers to

secure access to items in museums for legitimate purposes. Some respondents expressed feelings that museums may be overly restrictive regarding data sharing because of this.

In summary, respondents tended to agree that sharing the model and underlying data is important in communicating the results of their research to their readers. Respondents also tended to agree that current infrastructure to support and facilitate that sharing is inadequate.

Conclusion

It is evident from the findings that there is a role for libraries in supporting 3-D data sharing. Researchers are seeking out opportunities for standards development around issues that can be addressed collaboratively. One respondent said, on the topic of future work, “There’s no one person or one group of people that can make it happen, it really has to be the community.”

It is clear from the data that researchers are dissatisfied with infrastructure for sharing 3-D models, which they largely consider very important in communicating results. There are opportunities for libraries to explore more robust support for sharing 3-D models. Understanding the needs and expectations of the researchers we support and the research output that becomes part of our collections is central to the mission of research libraries. This work is just the beginning in terms of assessing the evolving role of 3-D modeling in research communities.

Researchers face obstacles in sharing and reusing 3-D output that include replicability, lack of documentation, and changeable publisher policies on rendering 3-D models in articles. Complex negotiations with communities who have rights to objects often result in restrictions on sharing data, as does a lack of guidance about ownership and copyright. This is an area that could benefit from enhanced support from library communi-

ties that work to raise awareness around issues of intellectual property, licensing, and authors’ rights.

The wide range of requirements participants shared for reuse of 3-D objects has significant implications for library sharing and publishing. Metadata standards and repository design can benefit from information related to reuse requirements. A thorough understanding of creation and reuse workflows ensures that library 3-D collections are relevant and accessible for use in research. Though some participants did not need technical or provenance data in order to reuse 3-D objects, excluding this information from library collections would be prohibitive for some researchers.

More work is needed to investigate the behavior of researchers who reuse data, approached as a distinct community from those who create data. Research communities are increasingly seeking platforms that enable publishing, sharing, and long-term access to digital 3-D models and accompanying data. Because of their stewardship, research, and access mission, libraries and other memory institutions are uniquely suited to offer services to support these needs. The findings reveal many areas of researcher need for 3-D support that intersect with existing library services: repository development, digitization, copyright consultation, publishing, collection development, metadata standardization, and data management. These results not only demonstrate the critical role of libraries in facilitating access to 3-D research objects, but they also provide a framework for developing services to support 3-D research output.

Funding

This study was funded by the Alfred P. Sloan Foundation #G-2014-13746 and the National Science Foundation NSF ACI #1349002 as part of a Research Data Alliance Data Share Fellowship Program. 