3D Research Data Curation Framework (3DFrame): Understanding 3D Data Creation, Analysis, & Preservation Practices across Disciplines

Three-dimensional, or *3D* data creation methods and virtual reality (VR)-based visualization techniques are increasingly common in interdisciplinary research and in the classroom. The 3D Research Data Curation Framework (3DFrame) project is a 3-year, $599,816 "Research in Service to Practice" project led by Dr. Zack Lischer-Katz (Asst Professor, UArizona, School of Information) and Matt Cook (Digital Scholarship Program Mgr, Harvard Library) that will study how researchers use 3D data, to align existing 3D data creation and curation methods with FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles (Wilkinson, et al., 2016) and develop methods for making 3D data usable with immersive visualization (e.g., VR) and other emerging analytic techniques. Using qualitative methods and emerging technologies, the PIs will develop a digital curation framework and publish an interactive toolkit that will enable academic libraries and other institutions to support the creation, curation, analysis, publication, and use of 3D data in immersive viewing environments for the benefit of researchers and instructors across disciplines. Public libraries will also benefit, because scholarly 3D sources can be brought into public libraries for K-12 instruction and community engagement (e.g., San Jose (CA) public Library's VR program and the XR Libraries Initiative) and treated as primary sources (e.g., Temple University Library's Virtual Blockson Project). This project is aligned with IMLS Program Goal 3 and Objective 3.2.

A. Project Justification

Researchers and instructors have come to rely on 3D data as both primary source material for original research and critical, experiential additions to remote teaching curricula.[1] However, demand for such activities and outputs presents a major challenge to existing digital curation frameworks for the academic libraries and other supporting institutions that are being asked to curate, preserve, and provide access to 3D data and support downstream visualization via VR and other immersive media (Cook, et al., 2019; Grayburn, et al., 2019; Hall, et al., 2019; Hardesty, et al., 2020; Lischer-Katz, et al., 2019; Moore, et al., 2022[2]). While 3D data formats share many of the same challenges as other complex data types (e.g., digital images and video) - including, large file sizes, proprietary formats, interdependent file structures, and critical post-processing decision-making that shape the integrity of the data - they also have unique characteristics that require rethinking existing digital curation approaches (Moore, et al., 2022). As the recent book *3D Data Creation to Curation: Community Standards for 3D Data Preservation* notes, "the features that necessitate treating preservation of 3D data differently emerge from the co-occurence of multiple data types [...] and multiple intents for end use (e.g., use to create fixed 2D visualizations, use to produce physical objects, use to allow interactive inspection, or use to produce a virtual experience)" (Moore, et al., 2022, p. 298). The distinctive characteristics of 3D data require digital curators to look more closely at the creation and reuse phases of the 3D data lifecycle and more fully consider the needs of creators and users of 3D data who may want to access or distribute these contents using virtual environments with immersive interfaces (i.e. "metaverses").

Researchers are producing high-resolution 3D datasets in a number of disciplinary contexts ranging from cultural heritage to STEM fields. From the perspective of libraries and other cultural heritage institutions, 3D contents are valuable for (at least) three reasons. First, accurately digitized, natural phenomena (i.e. physically extended objects of study) can function as primary source materials and facilitate engagement, similar to other document formats (Buckland 1991; Dale 1969). That is, researchers can study, measure, or otherwise analyze 3D surrogates just as they study objects "in the field". This is due in part because these surrogates more faithfully and comprehensively represent relevant physical features - like dimensionality, volume, texture, and scale - than do 2D images or textual descriptions. The value of 3D as a digital proxy for scholarly collections has been further demonstrated by the COVID-19 pandemic, which has increased the

[1] See Supportingdoc4.pdf for letters of support that demonstrate the range of concerns for the use of 3D and VR.
[2] See Supportingdoc1.pdf for full references.

demand for remote access to collections. When access to physical collections is limited, it is important that adequate creation and curation practices are used to ensure that 3D data are interoperable and reusable. Second, 3D datasets are useful for accessing collections of primary source material that are far from the researcher and difficult to travel to or are subject to access restrictions from custodial institutions. Finally, representing research findings using 3D methods supports different kinds of cognition and facilitates communication with non-experts and/or remote collaborators (Boyer 2016; Limp et al. 2011; Silvestri et al. 2010; Kingsley et al. 2019).

Although a handful of recent repository development efforts - including Morphosouce, based at Duke University, and Smithsonian 3D Collections - support ingestion, metadata customization, and online viewing of diverse 3D collections - a comprehensive, multidisciplinary understanding of the specific methods of initial 3D data creation and downstream analysis activities necessary for building digital curation models has yet to be developed. In order to maintain provenance and data quality, information institutions need digital curation models built on empirical analysis of the practices of 3D creators, in conjunction with downstream curation and access needs that make use of immersive interfaces.

Engaging with 3D assets in immersive viewing (i.e. AR/VR) environments provides additional benefits for researchers, especially for those working in disciplines "whose primary dimensions are spatial" (Donalek, et al., 2014), such as geology, atmospheric science, transportation, engineering, archaeology, medicine, chemistry, architecture, aerospace, and more (Acevedo, et al., 2001; Giacalone, et al., 2019; Helbig, et al., 2014; Janiszewski, 2020; Kingsley, et al., 2019; Milovanovic, et al., 2019; Seth, Vance, and Oliver, 2011; Uhr, et al., 2019; Safi, Chung, and Pradhan, 2019). The integrated use of 3D data and VR-based visualizations depends on the careful alignment of high-bandwidth visual interfaces (head mounted displays and tracking systems) with digital information sources (e.g., captured 3D data) representing spatially relevant specimens, objects, and artifacts. Accurately represented in an immersive viewing environment that supports embodied navigation and observation, researchers and students can view digital primary sources and interact with them as they would it in the physical world (Forsberg, et al., 2008; Gibson, 2002; Ragan, et al., 2012; Silvestri, et al., 2010). Conceived as a digital return to "the field" or a simulation of first-hand observations, with all the benefits afforded the embodied investigator therein, the academic implications of 3D/VR are profound, with the added benefit of allowing physically impossible analyses (Damerow, 2016; Van Dam, Laidlaw, and Simpson, 2002; Zhao, et al., 2019).

Prior to the release of affordable 3D processing software and VR equipment (taking place over the last eight years) the use of 3D and VR for research and instruction had been restricted to specialized labs and experimental classrooms (Dede, Jacobson, and Richards, 2017; Greene and Groenendyk, 2019; Jang, et al., 2017; Lischer-Katz, Cook, and Boulden, 2018; Mills, 2020; Pellas, et al., 2021; Portman, Natapov, and Fisher-Gewirtzman, 2015; Trelease and Rosset, 2008). Empirical research on VR during this period identified generalizable cognitive benefits (e.g., performance on spatial tasks related to search, comparison, and measurement) of VR under controlled conditions (Ragan et al. 2012; Laha, B., Bowman, D.A. and Socha, J.J., 2014; Forsberg, Katzourin, Wharton, and Slater, 2008). Integration of 3D creation and curation with immersive visualization techniques requires a careful alignment of pressing instructional needs and the high-quality output, metadata, and provenance demanded by researchers working in specific disciplines, especially those concerned about the reproducibility of measurements and other analyses across viewing software and hardware environments (Limp, et al., 2011; Boyer, 2016). Therefore, 3D data creation and immersive analysis methods must be designed to produce interoperable and reusable research outputs that are compatible with advanced visualization equipment and can be easily ingested into existing data management workflows and preservation repositories. In order to maintain provenance and ensure FAIR data quality for these research and instructional applications, libraries and other institutions need digital curation and service/support models that can simultaneously support the needs of researchers to develop new knowledge and the needs of librarians and instructors who want to curate, preserve, and use the content for teaching and learning.

This study will employ a research approach based on qualitative inquiry, in the tradition of the sociology of scientific knowledge (SSK) to answer the following research questions, which are aligned with current grand challenges in information studies in terms of research data management, reproducibility, and ensuring widespread access and usability of scholarly outputs:

Research Question 1: How do researchers evaluate and document the quality of research data throughout their creation workflows and the 3D data lifecycle?

Research Question 2: How do researchers use visual and immersive technologies to analyze their 3D data?

Research Question 3: How can institutions and curation professionals best support 3D data creation, analysis, and curation workflows?

Research Question 4: How can emerging technologies be used to support 3D data analysis, curation, access, and reuse?

Data gathered to address these research questions will allow the PIs to develop and deliver practical resources for information professionals, including: 1) Service/curation models (including roles and responsibilities) for libraries to support digital curation of 3D research data through their lifecycle; 2) evaluative metrics for documenting the quality of 3D research data; and 3) reproducible workflows for creating, evaluating, visualizing, and preserving 3D data.

The PIs will build upon findings of their previous, IMLS-sponsored project, *LIB3DVR*, which identified a pressing need among researchers, 3D creators, and library professionals for new digital curation models for 3D data formats (Hall, et al., 2019; Hardesty, et al., 2020; Lischer-Katz, et al., 2019). PIs have published extensively on the use of 3D/VR in academic contexts (e.g., Bozorgi and Lischer-Katz, 2020; Cook, et al., 2019; Cook and Lischer-Katz, 2019; Lischer-Katz, 2020; Lischer-Katz, Cook, and Boulden, 2018).

Project outcomes will support curation, research, and teaching in academic libraries, as well as produce resources for bringing scholarly materials into public and other libraries. The PIs have expertise that makes them exceptionally well positioned to advance this work: Lischer-Katz is co-editor of the Council on Library and Information Resources report, *3D/VR in the Academic Library*, and has used ethnographic methods to study how media preservationists construct knowledge through their work in preservation labs (Lischer-Katz, 2017; 2019; 2022); and Cook facilitates digital scholarship activities at Harvard Library, founded the emerging technologies unit at University of Oklahoma Libraries in 2015, and has published research on library innovation (Cook and Martens, 2019), immersive browsing (Cook 2018; Cook and Grime, 2021), and curricular integrations of 3D/VR technologies across disciplines (Cook and Lischer-Katz, 2019; Pober and Cook, 2019; Qin, Cook, and Courtney, 2021).

Critical guidance and context will be provided to the PIs by the project's diverse stakeholder advisory board, which includes digital curators from both sponsoring institutions; data repository (including 3D repository) managers; early career librarians; museum directors; faculty researchers; visualization and information technology experts; and private sector technologists representing major headset manufacturers. The advisory board members will provide their expertise for the design and dissemination of the products of this project (see Supportingdoc2.pdf for a list of committed advisory board members).

Innovations and Prior Contributions

Technologists and information professionals have begun engaging in the interdisciplinary study of the cognitive benefits of interacting with 3D objects in immersive environments (Donalek, et al., 2014; Kersten-Oertel, Chen, and Collins, 2013; Laha, Bowman, and Socha, 2014). Because existing research has been discipline and technology specific, the comprehensive, multidisciplinary understanding of the specific methods of 3D data creation and VR-based immersive analysis necessary for building digital curation models has yet to be developed.

To address these gaps identified in the literature and extend the generalized application of these tools and techniques to the creation and immersive viewing stages, we will study the work of current 3D data practitioners whose objects-of-study range in size from the very small to landscape-sized (Kelley and Wood, 2018; Publications Office of the European Union, 2022; Moore, et al., 2022). This scale-based approach guarantees that a mixture of scanning techniques and technologies are documented, including expensive, highly precise and accurate computed tomography (e.g. MicroCT), to structured light scanning devices, to versatile drone-based photogrammetric methods that are increasingly deployed for gathering cultural heritage and other spatially relevant data sets in the field (Alliez, et al., 2017; Pfarr-Harfst, 2016; Ropero-Miller, Bailey and Bushman, 2016; Soler, Melero, and Luzón, 2017; Gonzalez et al., 2023). Studying this continuum of scales and techniques across a range of disciplinary and technical methodologies will help the project team to identify common and divergent practices across research cultures.

Data gathered during the observation of 3D creators working in research sites will help the PIs develop a model of a rigorous 3D data pipeline that incorporates findings from international efforts to standardize the production of 3D objects (Publications Office of the European Union, 2022). This model will support the management of provenance, digital chain-of-evidence, tracking 3D data as it moves from practitioner research sites; to processing decision-making stages; to downstream users and curators via the web (for classroom use) and high-fidelity VR viewing systems; to hosting and preservation platforms such as the Harvard Dataverse, Open Science Framework, and Morphosource (Boyer, et al., 2016; Foster and Deardorff, 2017; King, 2007).

The project will also demonstrate exploratory methods for curating, analyzing and disseminating 3D outputs using virtual and augmented reality interfaces. A shared, immersive viewing environment ("metaverses") will be deployed online where PIs can host virtual events, to train practitioners in 3D data analysis techniques. An immersive showcase will demonstrate research applications of the toolkit. Referred to as "Immersive Analytics" in the literature, these emerging methods of engaging with and interpreting spatially-oriented data (e.g., point-to-point and volumetric measurement, object density, multi-spectral data overlays, and transparency and cutting-plane views, etc.) can help research assistants or data curators to evaluate data quality and aid in the detection of errors in the outputs before they reach the researcher or other downstream users or repositories (i.e., "VR-based quality control") and to make scanned source material more useful earlier in the research process (Skarbez, et al., 2019;). Alongside users, the project team will evaluate the usefulness of these and other emergent tools *in situ* in order to improve support for the 3D research data curation lifecycle. These techniques will be investigated to address gaps and points of breakdown identified through the qualitative data collection and analysis process.

Current activities in libraries and archives demonstrate the level of concern of information professionals for developing digital curation frameworks that support working with 3D data. For instance, a special CLIR report on 3D/VR curation, *3D/VR in the Academic Library*, and the recent publication by the Association for College and Research Libraries (ACRL) of a guide for working with 3D data in academic libraries and digital repositories by the Community Standards for 3D Data Preservation group, *3D Data Creation to Curation: Building Community Standards for 3D Data Preservation*, shows a perceived need for best practices, guidelines, and digital curation models (CS3DP, 2022). 3DFrame will deploy these innovations at a time of national need (i.e. the COVID-era) by studying the production, use, and preservation of 3D assets in the field, and thereby generating practical curation models and recommendations that leverage prior, theory-driven efforts and FAIR data principles.

Impact on Current Practice. The findings of this project will benefit multiple target communities: researchers who work with 3D data, information professionals who support the work of researchers who work with 3D data, educators who want to use 3D data in their teaching at K-12 and higher education levels, and public librarians who are making emerging technologies publicly available in their institutions. Findings of the LIB3DVR project (see "Supportingdoc5.pdf") and examples from the recent CS3DP book (CS3DP, 2022) have shown that guidelines, standards, and best practices for creating and curating 3D data are still being developed and are not widely adopted yet, and digital curators and libraries are still grappling with how to support the 3D data lifecycle or evaluate the quality of their data and use it in downstream visualization applications. The unexpected transition during the early phases of the COVID19 Pandemic to remote teaching and learning has exacerbated this deficit, as have highly variable 3D data creation practices, which now span a range of scales and production techniques, and require novel and sophisticated approaches for establishing guidelines. In particular, by collecting empirical data on the digital curation practices of researchers, this research hopes to identify those aspects of 3D creation that can be standardized or codified as general guidelines and protocols, and those that need to be open to local articulation. The parallel evaluation of 3D outputs by the PI team using advanced, immersive tools and methods, will ensure a practical toolkit, which will include technological recommendations as well as protocols for implementing quality control and error detection systems onsite.

Tensions between global standards and local practice have been widely noted in research in the STS and the sociology of standards (Bowker and Star, 1999; Lampland and Star, 2009). This project aims to ameliorate these tensions by emphasizing local and discipline-specific practice, while at the same time identifying generalizable guidelines that can help researchers, digital curators, and librarians alike manage, preserve, and provide access to this emergent form of data. By ensuring that 3D data conform to FAIR data principles in ways that fully consider the specificity and situatedness of creation practices, librarians and educators will be able to more easily access and integrate 3D data into educational programs, including for use in VR, providing an emergent platform that promotes immersive and engaging access to 3D content for a myriad of users.

B. Project Work Plan

Overview

To address the project's research questions, this project will use a comparative, multi-case study of 6 sites that are currently creating and analyzing 3D data. The PIs have established relationships with 4 institutions with research labs actively producing 3D data (Harvard, UArizona, UC-San Diego, and UMass-Amherst).[3] Partners at these institutions work with zoological specimens, mineralogical and archaeological specimens; and historical buildings and landscapes. The project team is in conversations with Pima County Public Library as an additional site, who has expressed interest in the project. One additional research site will be recruited to represent public libraries and smaller cultural heritage institutions (e.g. local and/or regional museums), through the guidance of the project's stakeholder advisory board. **Phase 1** of this 3-year project will consist of participant recruitment and data generation. PIs will travel to research sites to generate data and document participants' workflows, through participant-observation methods and interviews with 3D creators. The PIs will build computer workstations that will emulate the participants' working environments, enabling close analysis of how participants use 3D software and hardware for creation and analysis. In **Phase 2,** data will be transcribed and coded following grounded theory methods and guided by assumptions drawn from Science and Technology Studies (STS). The PIs will develop case studies and cross-case analyses, which will be used to develop data curation models, quality control criteria, and service models for libraries. In **Phase 3,** findings will be disseminated through conferences, publications, an online toolkit, and an immersive, multi-user online exhibit (i.e. an educational "metaverse") with VR support.

Project Timeline

Years 1 and 2 of the project will focus on collecting data from the five research sites selected for the study. The PI and Co-PI will plan and execute all research efforts with the support of two students to be hired at the start of the project: a graduate research assistant (GRA) stationed at UArizona, and an undergraduate research assistant (URA) located at Harvard Library. These student research assistants will be trained by the PI and Co-PI in methods of qualitative data collection analysis and the use of 3D scanning and virtual reality viewing equipment. The PI and Co-PI will design the research instruments and lead the research sessions at each site, including interviews and observation. The GRA and URA will be trained to assist in data collection, transcription, and initial data analysis. The PI and Co-PI will travel together to visit each site for 1 week to collect observational data, which will consist of a total of 50 hours of structured observation time (2 hours per

[3] See Supportingdoc3.pdf for a list of sites committed to participating in the study and their lab managers' names.

participant), and a total of 50 hours of site visit time (based on approximately two hours per day for five days at each of the five sites). Interviews and follow ups will be conducted over the phone. Either the PI or the Co-PI (plus relevant student research assistants) will conduct follow up visits to each site depending on where the site is located for a total of 50 hours of follow up site visit time (based on approximately two hours per day for five days at each of the five sites). Following the conventions of qualitative analysis, the data will be analyzed as it is collected. The GRA and URA will help to transcribe and begin data analysis. The PI and Co-PI will also begin analyzing and coding the transcribed data and begin constructing models of each site's main data curation workflows. Conference papers will be written based on preliminary findings and submitted to the ASIS&T Annual Meeting, the iConference, and the DLF Forum. In Year 3, the first half of the year will consist of student research assistants working with the PI and Co-PI to complete data analysis, as well as developing and testing techniques for error detection using VR technologies and aligning 3D the documented creation and curation workflows with FAIR data principles. From these findings, the PI and Co-PI will begin drafting publications and multimedia deliverables (e.g. a demo metaverse) to begin sharing the findings with practitioner and academic communities. In the remainder of year 3, the project team will develop the toolkit and will present papers and elements of the toolkit at conferences to disseminate research findings to broad audiences. Feedback on the toolkit and the training metaverse will be sought at conferences and from the advisory board. The final version of the toolkit will be promoted across listservs, social media, and through professional organizations. See "**Scheduleofcompletion.pdf**," for detailed project timelines for each year.

Theoretical Framework

The proposed research draws on a constructivist epistemology that assumes that "individuals are seen as actively constructing an understanding of their worlds, heavily influenced by the social world(s) in which they are operating" (Bates, 2006, p. 11). Research will be guided by the assumptions of qualitative inquiry and perspectives drawn from research in STS.Qualitative inquiry assumes that the generation and analysis of qualitative data requires *empathic identification*, such that "to understand the meaning of human action requires grasping the subjective consciousness or intent of the actor from the inside," which involves "trying to figure out (both by observing and by conversing) what the actors think they are up to" (Schwandt, 2000, p. 192). Our framework also draws on STS research in the tradition of the sociology of scientific knowledge (SSK), developed by Bruno Latour (Latour and Woolgar, 2013), Steve Woolgar (1988), and Karin Knorr Cetina (2013). These approaches question the understanding of scientific work, highlighting the ways in which the actual work of science in the lab deviates from what scientists say that they do, requiring study of actual lab-based practices. Studies in STS have looked at scientific labs and other sites of research, including the visual practices of meteorologists working in a weather center (Daipha, 2010; 2012; 2015) and work on a scientific research vessel (Goodwin, 1994). This approach has been extended by Lischer-Katz (2017; 2019; 2022) to the study of digitization labs and the information practices of media preservationists.

Methods

In order to develop the 3D data curation framework and toolkit, the project will use a comparative, multi-case study approach, studying six individual research sites that currently create and/or analyze 3D data as a central part of their research process. This will involve looking closely at the processes of 3D data creation, how the data is analyzed, their documentation techniques, classification systems, and metadata schemas, and how they configure their computer hardware and software to facilitate their data workflows. The research team will analyze samples of 3D files at different points in their workflow and will emulate participants' workstations to better understand their technical workflows and to identify gaps or moments of breakdown in the process. The project team will evaluate the usefulness of new 3D data curation tools, such as VR-based quality control and error detection techniques to address the identified gaps and breakdowns. This will help the project team develop error detection protocols and technology recommendations that will be included in the toolkit for the practical benefit of stakeholders. The project will also recruit two smaller institutions, such as public libraries, which will allow the project team to study how 3D and VR are being used beyond academic contexts and find ways in which academic data creation can support education in other contexts.

Site Selection. The unit of analysis will be at the level of the 3D creation lab, or "site", which typically consists of small groups of technical staff (technicians, librarians, graduate students, etc.) working to assist faculty researchers in the creation of their 3D research data. Labs often support single long-term projects but can also support multiple projects. Criteria for selecting sites include: 1) sites must currently be creating 3D research data; 2) sites must be practically accessible to the researchers; and 3) site managers must be comfortable with having researchers observe their spaces. In order to collect data from a diversity of approaches, labs will be recruited from the Southwest, the Northeast, and the Central Plains. This has the benefit of giving insight into projects across the country, and because the project PIs are located in the Southwest and the Northeast, it also will require less travel to visit some of the research sites.

Because the proposed toolkit is intended to serve multiple research areas and 3D creation techniques, site selection will also be carried out to ensure that multiple disciplines (natural sciences and humanities), scales of production (small and large objects being scanned) and modes of 3D creation (photogrammetry, computed tomography [CT] scan, and LiDAR 3D creation techniques) are captured in the study. This will support cross-case analysis of data, which will help the PIs to derive common attributes of all workflows and points of divergence and idiosyncrasy. PIs have secured commitments from labs/projects at Harvard, UArizona, and UMass-Amherst. One additional site will be recruited from smaller institutions, so that the project findings will benefit institutions of varying sizes and diverse stakeholder groups. The guidance of the project's stakeholder advisory board also will help ensure that the research is conducted with diverse input (see Supportingdoc3.pdf and Supportingdoc4.pdf for advisory board information). Participants will be recruited at each site through the site's manager (i.e., the person who has institutional authority over activities at the site).

Participant Recruitment. It is expected that up to 5 participants will be recruited from each site, with an expected maximum enrollment of 40 participants. If individuals choose to participate, PIs will provide them with informed consent forms to fill out and set up site visits for interviews and observations during the 3D production process. Interviews may be conducted via videoconferencing if necessary, but observations need to be conducted in-person in order to observe the process of 3D data creation *in situ*.

Data Collection. In years 1 and 2 of the project, the project team will travel to research sites to collect data by conducting semi-structured interviews, participant observation, and through collecting locally produced documents related to the 3D creation process. Semi-structured interviews will generate data that will provide insight about participants' 3D creation workflows, as well as their epistemic practices of 3D analysis and evaluation of 3D data quality. Interviews are expected to run from up to 1 hour per participant and will be audio recorded and transcribed by the project's graduate and undergraduate research assistants (for a total of 40 hours of interview recordings). Observation will be augmented with a technique called "think-aloud" that encourages participants to put into words what may typically be unconscious, tacit, or otherwise hidden in their embodied practices (Ericsson and Simon, 1993). Each participant will also be observed for two hours, as they work through their 3D process. Participant-approved video recording will be used to capture the participants' activities throughout the participant observation phase. These recordings will capture audio and images that will be analyzed by the researchers, and they also serve as a document that can help participants further describe their practices while watching the video recordings, which can help clarify the observations and analysis of the researchers. Research data and drafts of data analysis write-ups will be shared with participants throughout the data collection and analysis phases and their feedback will help shape the toolkit. Site visits will also be conducted, in which project team members will observe the overall functioning and culture of each lab/project work space. Site visits will result in field notes and photos, which help support the analysis of the interview and observational data.

Data Analysis. Both the qualitative data collected from participants and the 3D datasets created by participants will be analyzed as part of the data analysis process, with qualitative methods taking place during years 1 and 2 of the project feeding into year 3 when the 3D output analysis activities designed to address gaps and points of breakdown along the 3D data pipeline as observed through cross-case comparison.

Computer-based, qualitative analysis software, MAXQDA software will be used to organize and code all collected documents, interview transcripts, and video recordings. Line-by-line data coding will be conducted following grounded theory procedures to draw out emergent codes from the data. Creswell (2014) defines grounded theory as “a design of inquiry from sociology in which the researcher derives a general, abstract theory of a process, action, or interaction grounded in the views of participants. This process involves using multiple stages of data collection and the refinement and interrelationship of categories of information (Charmaz, 2006; Corbin and Strauss, 2007)” (p. 14). Employing open coding, axial coding, and memo-writing enables a systemic approach to inductively analyzing the data without committing to the full grounded theory regime. Because of the emergent nature of a qualitative research design, data analysis goes hand-in-hand with data collection (Creswell, 2014), thus data analysis will begin as data is still being collected. In August 2023, we will obtain UArizona Institutional Review Board (IRB) approval for all research protocols and instruments prior to participant recruitment and data collection.

Evaluation of 3D Data Analysis Tools. Based on the findings of the data analysis process, PIs will evaluate the efficacy of emerging 3D data analysis tools to address gaps and breakdowns in existing 3D curation workflows. The project will evaluate 3D processes with VR-enabled PC workstations, using samples of 3D data from different research stages to develop quality control protocols and associated technology recommendations for inclusion in the 3DFrame toolkit. Because users can rely on “A flick of the eye or a turn of the head…to consult a different data source,” (Andrews, Endert, and North, 2010) the combination of both increased viewing space and head (and body) tracking allows users to offload cognitive workload – especially visuo-spatial working memory – to the system itself and thereby increase the efficiency of analysis tasks (Anderson, et al., 2016; Batch, et al., 2019; Lages and Bowman, 2018; Pirolli and Card, 2005; Whitlock, Smart, and Szafir, 2020).

The research team will apply immersive analytics to the 3D data collected from participants’ workflows to identify and address potential workflow gaps that may hinder FAIR data principles when 3D outputs are analyzed with VR. Immersive viewing tools will be used to analyze 3D data for errors (e.g, mesh artifacts and holes, color inconsistencies, scale accuracy issues, etc.), using new and emerging techniques documented in the literature (Kluge, et al., 2019; Prouzeau, et al., 2019; Zhao, et al., 2019). The project team will develop procedures and technology recommendations that will help stakeholders construct their own VR-based quality control workstations for the detection and correction for errors in their 3D datasets (Laha, Bowman, Socha, 2014; Whitlock, Smart, Szafir, 2020; Bharathan, et al., 2013; Brunken, et al., 2003; LaViola, et al., 2017).

Validation Strategies. The epistemological assumptions of interpretive qualitative research methods - particularly the assumption that reality is constituted through intersubjective, conventionalized uses of language - make it problematic to assess the validity and quality of the findings of interpretive research with quantitative techniques. Angen (2000) argues, “interpretive research is a chain of interpretations that must be documented in order for others to judge the trustworthiness of the meanings arrived at in the end (Nielsen, 1995)” (p. 390). To that end, this project employs several validation strategies: 1) It generates multiple forms of data in order to understand the practices of 3D data creation from multiple perspectives. 2) This research includes “member checking” (Creswell, 2014) as a strategy for bringing major themes from the analysis back to the participants and encouraging them to respond, to let the research team know how closely their interpretations fit with their lived experiences. 3) We will include extensive quotes from participants and rich descriptions of their work activities to provide a complete picture of the social context and the actions and statements of participants. By providing representative extracts of the data upon which we base our claims, readers will be encouraged to make their own interpretations of the data.

Choice of methodology and appropriateness to address research questions. The multiple challenges of studying the practices of researchers have been identified within science and technology studies (STS) and related fields. Collins (2001), for example, showed how knowledge can be tacit and invisible even to members of the same lab. This suggests that qualitative methods that incorporate data generation in the form of participant interviews and observations of their work can yield a more complete picture of what is happening in the contexts of 3D data creation and use. Studying the work of scientists in their labs or other research sites is an established approach in the sociology of scientific knowledge (SSK), including early work by Garfinkel, Lynch, and Livingstone (1981); Knorr-Cetina (1981; 1999); Latour and Woolgar (1981); Law and Williams (1982); and Pinch (1985), among others. Processes and products of visual representation play important roles in scientific inquiry (Pauwels, 2006). Researchers in information studies have studied the work of scientists in order to understand their information needs, seeking, and use, as well as their practices of producing and managing data, as well as in order to understand what are termed "knowledge infrastructures," which refer to configurations of social, technical, and documentary regimes that support the production of scientific knowledge. Digital curation can be seen as an important component of knowledge infrastructure, which has been broadly investigated by Borgman (2015), Borgman, et al. (2015), and Darch, et al. (2015) using qualitative research methods. Other studies have sought to link researcher behavior to library services and resources by looking at how university libraries support the data curation practices of researchers (Darch, et al., 2020; 2021). Qualitative methods have shown promise for supporting inquiry into research practices, and their research outputs support the work of libraries and other institutions.

Translation and Dissemination Plan. We plan to disseminate project findings of target communities in several ways. First, project participants will have access to our research findings as we produce them, through member checking and conversations. Second, we will share our findings with information researchers and professionals through a series of research articles and conference presentations in both researcher-oriented forums such as iConference, International Conference on Preservation (iPres), and the Association of Information Science & Technology (ASIS&T) Annual Meeting, as well as in practitioner-oriented conferences such as the Digital Library Federation (DLF) Forum. Findings will be published in academic and practitioner-oriented journals, with a focus on open access journals, including *College & Research Libraries*, *Library and Information Science Research*, *Information Technology and Libraries*, *Journal of Data Science and Information Science*, and *Information Research*. Third, the resultant toolkit will synthesize the practical implications of our findings in an informative set of guidelines, workflow diagrams, glossaries, data curation models, and weblinks to additional 3D data curation resources and guidelines (e.g., links to open source software and recommended metadata schemes, 3D file test kits, configurations of equipment, software recommendations, etc.) that will be hosted on the project's website and will serve as a reference guide for researchers and information professionals. The toolkit will be disseminated through the listservs of professional organizations, such as ASIS&T and DLF, and further promoted through conference presentations. De-identified qualitative research data (following IRB guidance) will be deposited with the University of Arizona Research Data Repository (https://arizona.figshare.com/). The toolkit will also be demonstrated through an immersive online training environment populated with sample 3D models alongside interactive media (links, videos, photos) documenting best practices in the field. PIs will also make use of open data hosting platforms (OSF, Dataverse, GitHub, Morphosouce) to manage data and publish supplementary material and documentation for reference and reproducibility purposes and to encourage access to non-academic and underrepresented scholarly communities (Sullivan, DeHaven, and Mellor, 2019).

Project Evaluation and Reporting

Throughout the award period, the PI will provide annual reports to IMLS on project performance and financials. The impact of project outputs will be evaluated in several ways:

- The reports and journal articles will be evaluated based on downloads, citations counts, and altmetrics.
- The toolkit and training metaverse will be evaluated based on downloads, local integrations, and user access counts.
- Anonymous surveys will also be distributed to the users of the toolkit and training metaverse to evaluate the usability and usefulness of the project deliverables.

C. Diversity Plan

To ensure diverse perspectives and broad usability of the toolkit for underrepresented communities, the PIs have secured the support of experts with diverse backgrounds, status, and institutional affiliations for the stakeholder advisory board (see "Supportingdoc2.pdf"). This includes early career information professionals of color; women working with data repository and K12 educator communities; and non-faculty technologists, as well as faculty researchers and instructors representing national public and private institutions. Beyond informal consultations, this cohort of advisors will meet together with the PIs three times per year, for updates on the project status. During these formal meetings, the advisory board will be asked, specifically, to comment on the trajectory, and potential impact of 3DFrame project activities as it applies to BIPOC (Black, Indigenous, and people of color) and LGBTQ communities and associated stakeholder groups. Additional 3D creation sites will be recruited from smaller institutions, so that the project findings will benefit institutions of diverse sizes and stakeholder groups. UArizona is designated as a Hispanic Serving Institution (HSI) and has a long history of supporting Native American students, faculty, and communities. The project team will look to guidance from the various programs at the iSchool and across UArizona, such as the Knowledge River program (https://ischool.arizona.edu/knowledge-river) and the offices of Diversity and Inclusion (https://diversity.arizona.edu/) to ensure that research practices and the promotion of research findings benefit these local and national communities.

D. Project Results

Understanding how researchers are creating and curating their 3D data will advance knowledge and best practices for 3D data creation workflows and curation strategies for libraries and other collecting institutions across the U.S. The main aim of this project is to develop a digital curation framework and associated toolkit that will enable researchers, academic libraries and other institutions to support the creation, curation, analysis, and preservation/access for 3D data. The framework will consist of 1) digital curation models that define best practices for establishing workflows, assigning roles and responsibilities, and capturing metadata throughout the curation lifecycle; 2) qualitative and quantitative metrics for evaluating the quality of data; and 3) guidelines for ensuring datasets are reusable outside of the contexts of initial creation, including the use of immersive data analysis techniques and technology recommendations.

The framework will be packaged as a toolkit consisting of the guidelines and models produced through the project, as well as links to software, hardware, and additional digital resources necessary for implementing the recommendations of the framework. The toolkit will be exhibited in an online, immersive showcase, using a combination of sample 3D objects and associated multimedia. This VR-ready metaverse will give scholars and practitioners an interactive way to engage with the concepts and recommendations presented by the toolkit. The project team will work with a VR consultant to design metaverse modules.[4] In order to build knowledge about the practices of 3D data creators and synthesize guidelines, qualitative research methods and advanced visualization techniques will be used to address the project's research questions. The PIs will publish findings in top-tier LIS practitioner and research journals, and will create an open access toolkit (i.e. workflows, quality control procedures, etc.) that will provide guidance for library and information science professionals and aid in the design of 3D data repositories and curation tools. The toolkit will consist of a guide containing models and techniques for customizing digital curation approaches, following the Data Curation Profiles and the OSSArcFlow Projects.

Research outputs will be useful for other projects that are also developing guidelines, including the CS3DP project, which recently published a guidebook, including case studies, for supporting 3D data curation and preservation in libraries, but does not include empirically-grounded analysis of 3D creation practices gathered from the field. Members of 3D initiatives, experts on VR, library and museum practitioners, digital curators, and researchers have been recruited for the advisory board to ensure broad usability of project outputs. Further, the toolkit will include practical methods and recommended technology for impactful classroom deployment of 3D outputs, as well as quality control methods to ensure that early-stage products meet FAIR data standards prior to VR analysis and/or ingestion into institutional repositories, and future scholarly and pedagogical uses.

---

[4] See Supportingdoc6.pdf for a price quote and a letter of commitment from the vendor, Proximal VR, LLC.

IMLS NLG: 3D Frame Project

Schedule of Completion (Year 1: August 1, 2023 - July 31, 2024)

Zack Lischer-Katz (PI/Project Director); Matt Cook (Co-PI)

	Objectives	2023			2023-2024			2024			2024		
	Schedule of Completion	Aug	Sept	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr	May	June	July
1.0	**Project Planning & Startup**	■	■	■	■								
1.1	Refine research instruments (interview and observation protocols)	■	■	■									
1.2	Develop project protocol for IRB submission	■	■	■									
1.3	Hire graduate research assistant (GRA) for UA-iSchool site and undergraduate research assistant (URA) for Harvard Library site.	■	■	■									
1.4	First meeting of stakeholder advisory board		■										
2.0	**IRB submission and participant recruitment**			■									
2.1	PI and Co-PI: Submit protocol to IRB, train students in qualitative data collection methods.			■									
2.2	GRA and URA: Begin Participant recruitment at sites-of-study; renew CITI training for human subject certification.			■									
3.0	**Conduct Pilot Study at Site #1 and Refine Protocol**				■	■							
3.1	PI and GRA: Pilot research protocol at local U of Arizona site.				■	■							
3.2	PI and Co-PI refine protocol based on data collected from pilot.				■	■							
3.3	GRA and URA trained in qualitative data analysis methods by PI.				■	■							
3.4	Secod meeting of stakeholder advisory board							■					
4.0	**Data Collection and analysis at Site #2**						■	■	■				
4.1	Full project team travels to Site #2						■						
4.2	Follow up visits and calls (PI or Co-PI depending on location)							■	■				
4.3	Third meeting of stakeholder advisory board										■		
5.0	**Data Collection and analysis at Site #3**								■	■	■		
5.1	Full project team travels to Site #3									■			
5.2	Follow up visits and calls (PI or Co-PI depending on location)										■		
6.0	**Regroup and Planning: June - July 2024**											■	■
6.1	PI and Co-PI: Review data, organize research notes, and plan for year 2											■	■
6.2	PI and Co-PI reconcile accounts and submit annual report to IMLS											■	■

IMLS NLG: 3D Frame Project

Schedule of Completion (Year 2: August 1, 2024 - July 31, 2025)

Zack Lischer-Katz (PI/Project Director); Matt Cook (Co-PI)

	Objectives	**2024**			**2024-2025**			**2025**			**2025**		
	Schedule of Completion	Aug	Sept	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr	May	June	July
1.0	**Data Collection and analysis at Site #4**												
1.1	Full project team travels to Site #4												
1.2	Follow up visits and calls (PI or Co-PI)												
1.3	Project team travels to ASIS&T conference to present initial findings												
1.4	Fourth meeting of stakeholder advisory board												
2.0	**Data Collection and analysis at Site #5**												
2.1	Full project team travels to Site #5												
2.2	Follow up visits and calls (PI or Co-PI depending on location)												
2.3	PI and Co-PI travel to DLF Forum to present practical findings.												
3.0	**Data Collection and analysis at Site #6**												
3.1	Full project team travels to Site #6												
3.2	Follow up visits and calls (PI or Co-PI depending on location)												
3.3	Project team travels to iConference to present initial findings.												
3.4	Fifth meeting of stakeholder advisory board												
4.0	**Data Analysis**												
4.1	Full project team codes analyzes data, conducts cross-case analysis, writeups												
5.0	**Regroup and Planning for Next Year**												
5.1	Sixth meeting of stakeholder advisory board												
5.2	PI and Co-PI: Review data, organize research notes, and plan for upcoming activities for year 3												
5.3	PI and Co-PI reconcile accounts and submit annual report to IMLS												
5.4	Configuration and deployment of VR workstations for 3D data analysis												

IMLS NLG: 3D Frame Project

Schedule of Completion (Year 3: August 1, 2025 - July 31, 2026)

Zack Lischer-Katz (PI/Project Director); Matt Cook (Co-PI)

	Objectives	2025			2025-2026			2026			2026		
	Schedule of Completion	Aug	Sept	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr	May	June	July
1.0	**Data Analysis and Toolkit Outline**												
1.1	GRA and URA: Continue data analysis under guidance of PI and Co-PI												
1.2	PI and Co-PI: Based on findings, begin drafting research papers and outline toolkit components, including quality control protocols												
1.3	PI and Co-PI: Travel to DLF Forum to promote toolkit and get feedback on structure of outline.												
1.4	Seventh meeting of stakeholder advisory board												
2.0	**Toolkit Draft and Journal Publications**												
2.1	Project team will continue to develop toolkit and present it at conference for feedback, including iConference												
2.2	PI and Co-PI, with assistance of students begin drafting journal publications to disseminate research findings.												
2.3	Eight meeting of stakeholder advisory board												
3.0	**Finalize Toolkit and Disseminate**												
3.1	Project team will finalize the toolkit and begin to promote and share widely across listservs and social media.												
3.2	PI and Co-PI, with assistance of students will continue to draft articles and submit to professional journals in order to disseminate research findings an promote final toolkit.												
3.3	Ninth (final) meeting of stakeholder advisory board												
4.0	**Project wrap-up and Final Report for IMLS (PI and Co-PI)**												
4.1	PI and Co-PI Reconcile Accounts and Write up final report for IMLS												
4.2	PI Deposits Digital Products and Research Data with Repositories												

3D Research Data Curation Framework (3DFrame)
Digital Products Plan

Type: *What digital products will you create?*
The project will create digital resources in two forms: a series of journal articles published in open access journals and a toolkit, which will take the form of a website that will host project deliverables, including VR applications. Personal computers will be used to assemble the project files and create a website (using a CMS such as Wordpress or Drupal). Adobe Photoshop and Illustrator software will be used to create the images and diagrams to depict participant workflows and develop models for recommended specifications and protocols. Sample 3D files will be collected from participants and incorporated into the set of web-accessible resource files. The VR project will be developed with a VR consultant, who will author the VR project using an industry-standard 3D game development platform, such as Unity3D or Unreal, to make it easier to archive the project files and migrate the project to new platforms in the future. The published versions of the journal articles will take the form of PDF/A files that will be deposited with the UArizona Institutional Repository (https://repository.arizona.edu/), to enhance discovery and ensure long-term preservation. The toolkit will take the form of a website, which will consist of a set of HTML files (8-10), and related digital assets, which will include images (20-30 JPEGs), diagrams (8-10PNGs), & 3D test files for users to use with the workflows and recommendations of the toolkit (includes a set of 3D files [between 20-30] stored in a variety of formats, including X3D, PLY, DAE, OBJ, and other formats). The website will also host the VR products of the project, which will consist of VR packages for loading on current VR equipment (e.g., APK [Android Package Kit] files, for loading on Oculus Quest2 VR headsets).

Availability: *How will you make your digital products openly available (as appropriate)?*
Digital content will be made freely available via these methods: The articles published in open access journals will be made available through the University of Arizona Institutional repository to make the content more discoverable across the university and searchable by the public, librarians, and members of other academic institutions. The toolkit will take the form of a webpage with associated 3D resource files, workflow diagrams, images, and links to open source hardware and software sources across the Internet. It will be linked from the project website, which will be promoted through the University of Arizona iSchool webpage (http://ischool.arizona.edu), the PI's website (http://zacklischerkatz.com), the PI's Lab website (The co/lab - the Critical Archives & Curation Collaborative). The de-identified data will be deposited at UA Research Data Repositories: University of Arizona Campus Repository: (https://repository.arizona.edu/arizona/) and the University of Arizona Research Data Repository (https://arizona.figshare.com/). In addition, all of the project's open access resources will be promoted via social media, research and practitioner listservs (DLF, ASIS&T, JESSE, ALISE, etc.), and presentations at conferences (ASIS&T, iConference, DLF Forum, etc.).

Access: *What rights will you assert over your digital products, and what limitations, if any, will you place on their use? Will your products implicate privacy concerns or cultural sensitivities, and if so, how will you address them?*
The copyright of any products produced by the research team will remain with the authors, in accordance with Arizona Board of Regents (ABOR) "Intellectual Property Policy," #6-908, (https://public.azregents.edu/Policy%20Manual/6-908-Intellectual%20Property%20Policy.pdf), which states that faculty retain the copyright for scholarly works that they produce. To ensure the broadest sharing and reuse of the scholarly outputs of this project, the project team will prioritize publication in open access journals

and license all resources, websites, and toolkits under a Creative Commons "Attribution-ShareAlike 4.0 International" (CC BY-SA 4.0) license (https://creativecommons.org/licenses/by-sa/4.0/). A notice will appear at the bottom of resources shared. The PIs will grant IMLS a royalty-free, nonexclusive, and irrevocable right to reproduce, publish, or otherwise use the digital products and authorize others to reproduce, publish, or otherwise use the digital products. Any 3D models or VR products will be evaluated for any cultural sensitivity issues. In addition, examples of 3D models produced at research sites will be made available with the permission of those research sites under a CC BY-SA 4.0 license, as negotiated with the research sites for inclusion in the VR products of the project. As part of the research study, the PIs will be collecting qualitative data from research sites. Before research begins, we will be submitting the research protocol to the UArizona Institutional Review Board (IRB) for their approval. This will ensure that the privacy of our participants will be protected. Any data made publicly available will be de-identified. We will follow the policies of the UArizona Research Data Repository, which specifies "Data produced from activities that are classified as human subjects research can be deposited ONLY if (a) such data has been de-identified (or consent has been obtained to share identifiable information), (b) the consent form contains language allowing this sharing, and (c) the IRB has approved the sharing of the Data."

Sustainability: *How will you address the sustainability of your digital products?*
The project's digital products have different sustainability concerns, with the journal articles and textual forms of the toolkit having clearly defined long-term preservation guidelines. PDF/A files that will be deposited with the UArizona Institutional Repository (https://repository.arizona.edu/), to enhance discovery and ensure long-term preservation. PDF/A files are the standard for archiving textual content.

The toolkit, which consists of a website, text, images/diagrams, 3D models, and VR applications, requires more consideration to maintain over time, but the project team is committed to keeping the files accessible and renderable for at least the five years after project completion (at which point, new 3D and VR techniques and tools will be available. At that point, digital products will be appraised for long-term value and migrated to new 3D and VR platforms. The website can be archived as HTML files, or if the look and feel of the website is deemed important, a web archiving service, such as the Internet Archive's Archive-It, would be used. JPEG and PNG images are mature and stable formats that will likely still be renderable in the next 5-10 years, as there are currently no perceived risks to obsolescence. Standards for preserving and archiving 3D models are still developing, but the normalization of formats to open, well-documented formats, such as DAE and X3D files will be conducted to ensure 3D files are renderable in the future. Standards and best practices for archiving VR projects are also evolving. Conservationists at the Tate modern (citation needed) stress the importance of archiving the authoring files and related software versions to ensure that VR projects can be opened in the near future and migrated to the newest platforms. The VR products of the project will be archived VR packages for loading on current VR equipment (e.g., APK [Android Package Kit] files, for loading on Oculus Quest2 VR headsets), as well as all of the 3D project files (e.g., Unity3D project and asset files) so that the VR project can be opened up and migrated to the next VR platforms. All of the digital products of this project will be stored on the project team's OSF site, the project website, and at the end of the project, deposited at the UArizona Libraries - data services, which is committed to long-term bit-level preservation of deposited faculty research products. The PIs will need to handle any migration of the complex media formats (3D, VR, etc.) to new formats as needed, which would be an ongoing task of the PI and the co/lab (and potential for more research opportunities studying the long term preservation of these formats).

3D Research Data Curation Framework (3DFrame): Understanding 3D Data Creation, Analysis, and Preservation Practices Across Disciplines
Data Management Plan

This data management plan will be reviewed by the PI in consultation with the Co-PI quarterly during the data collection phases of the project. The methods and tools of data sharing between PI and Co-PI will be modified if they agree that it is necessary.

Intellectual Property Rights and Permissions

The copyright of any products produced by the research team will remain with the authors, in accordance with Arizona Board of Regents "Intellectual Property Policy,", which says that faculty retain the copyright for scholarly works that they produce. To ensure the broadest sharing and reuse of the scholarly outputs of this project, the project team will prioritize publication in open access journals and license all resources, websites, and toolkits under a Creative Commons “Attribution-ShareAlike 4.0 International” (CC BY-SA 4.0) license (https://creativecommons.org/licenses/by-sa/4.0/). A notice will appear at the bottom of resources shared.

Qualitative Data

The project will collect qualitative data from participants in the form of interview audio recordings and video recordings of site visits and observation/think aloud sessions; the transcripts of all audio or video data collected; documents related to the workflows of participants (with their permission); and field notes and research journals kept by the research team throughout the process of data collection, analysis, and drafting of publications. This data will be used to develop the workflows, guidelines, and suggestions for 3D creation and curation processes, which will be synthesized and made publicly available via journal articles, conference presentations, and a web-based toolkit. This research data will be collected at research sites during the first two years of the project. The data will be analyzed as it is collected and will be used to develop research articles and the toolkit in year 3. Human subjects data will be collected so review by UArizona's IRB will be required. The PI will apply for IRB approval during the start-up phase of the grant (August - October 2023) before the data collection phase begins. Some of this data could be sensitive as participants may reveal personal information about their work practices in their interviews that they may not want associated with them (or there may be risks to their professional careers). Therefore, recordings and transcripts will be de-identified and analyzed (each participant will be given a code, and the codebook will be deleted after analysis). Participants will complete informed consent forms, which will be collected & stored in the PIs locked cabinet in his office. A shared spreadsheet (hosted on Google Drive) will enable the project team to document the day and time of interviews, observations, and site visits using the de-identified codes assigned to participants. Metadata about files collected and file names will be organized in sheets in the same shared spreadsheet. Transcripts will be created from audio and video files collected from research sites. Cloud-hosted MAXQDA qualitative analysis software will be used to organize and code the data. This data will be password-protected & accessible by the project team. De-identified data will be exported as CSVs.

Data Formats

The data generated in this project will consist of qualitative research data, including: audio recordings and transcripts of interviews with participants (WAV files for the recordings, and .docx files for the transcripts) videos (MP4s) and digital images of workplace practices of 3D data creators (JPEGs). Observation notes and memos will also be kept by the project team in a shared in lab notebook stored on Open Science Framework. Total data collected is expected to be: Audio: ~57GB (45 participants, 2 hours of interviews per participant, CD quality audio: 16bit/44.1k,); Video: ~3.4TB (45 participants, 2 hours of recording per participant, h.264,1080p, 24fps), Images: ~3.0GB (6 research sites, 200 images at each site, 12 megapixels, JPEG [100%, 24bit]), and

Text: ~30MB (6 research sites, 100 pages of notes per site, .docx format). Sample 3D files will be collected from participants and incorporated into the set of web-accessible resource files. The qualitative data will be captured in the following formats: WAV, MP4, JPEG, PDF, and DOCX. The project will also use the following formats for creating digital products for the project: PDF/A, HTML, JPEG, PNG, 3D formats (X3D, PLY, DAE, OBJ, etc.), and VR applications (e.g, APK format).

Quality Control of Data and Digital Assets

The student research assistants will be responsible for generating PDF/A, JPEG, and PNG files, and for collecting 3D files from sites. Before files are made accessible to the public they will be checked by the PI or Co-PI. Quality control will involve checking that the correct files are included, and have the correct file names, complete metadata, and file extensions applied to each file. 3D file formats will be checked against checksums generated at the time of file creation. 3D file formats will also be validated for compliance to file format specifications using JHOVE software (https://jhove.openpreservation.org/).

Preservation

During the project, all of the project files will be stored via Open Science Framework (OSF) (https://osf.io/), which will ensure that project files are shared and made accessible to the project team and that versions of files can tracked and metadata for files can be managed. This is a free service. At project completion, all digital materials (except research data) will be deposited with the University of Arizona Institutional Repository (https://repository.arizona.edu/arizona/). De-identified research data will be deposited at the UArizona Research Data Repository (http://arizona.figshare.com). Funding for these repositories is provided by the University of Arizona Libraries, which has committed to the long-term preservation of faculty scholarly outputs. This is a free service provided to members of the UArizona research community.

Metadata

Descriptive metadata will be created for the dataset using the Dublin Core-based schema developed by the Open Archives Initiative Protocol for Metadata Harvesting: http://www.openarchives.org/OAI/2.0/oai_dc/ (which is used by the UArizona Research Data Repository). This information will be maintained on the OSF site during data collection and analysis. At the time of deposit, this metadata will be included with the de-identified data when it is deposited with the UArizona Research Data Repository. In addition, a README.TXT file will be included with the dataset at the end of data analysis, describing the project goals, methods of data collection and analysis, license information and project team member information for attribution, and the data fields included in the dataset. Metadata will be stored on the project's Open Science Framework (OSF) site. This will facilitate sharing metadata between project team members and enable the maintenance of metadata throughout the data collection and analysis phases of the project. At the completion of the project, the metadata will be included with the data deposited at the UArizona Research Data Repository. When the data is deposited, a DOI will be issued, which will help associate scholarly outputs back to the underlying data.

Access and Use

Data and research outputs will be made freely available via these methods: The articles published in open access journals will be made available through the University of Arizona Institutional repository to make the content more discoverable across the university and searchable by the public, librarians, and members of other academic institutions. The toolkit will take the form of a webpage with associated 3D resource files, workflow diagrams, images, and links to open source hardware and software sources across the Internet. The de-identified data will be deposited at UA Research Data Repo. All of these open access resources will be promoted via social media, research and practitioner listservs (DLF, ASIS&T, JESSE, ALISE, etc.), and presentations at conferences (ASIS&T, iConference, DLF Forum, etc.).

3D Research Data Curation Framework (3DFrame): Understanding 3D Data Creation, Analysis, & Preservation Practices across Disciplines
Organizational Profile

A. Organizational Description: The University of Arizona (https://www.arizona.edu/) is a public land-grant university in Tucson, Arizona, currently serving over 46,932 students in 19 colleges/schools. It is a Hispanic Serving Institution (HSI) located on the traditional homelands of the Tohono O'odham Nation, and was established in 1885 as the first university in the Arizona territory, and is currently classified as an "R1: Doctoral Universities - Very High Research Activity." Its official mission is to "continuously improve how we educate and innovate so we can lead the way in developing disruptive problem-solvers capable of tackling our greatest challenges" in order "to create a world where human potential is realized and we're all working together to create solutions to big problems, so that life in our communities, in Arizona and on our planet can thrive" (https://www.arizona.edu/purpose-values).

B. External Funding
    1. **Proposed Project:** N/A
    2. **Proposed Project Director:** N/A

C. External Partnerships: Harvard Library (https://library.harvard.edu/) will be a subrecipient of this grant award and partner on this project. Harvard Library is the umbrella organization for Harvard University's library system, serving its students, faculty, staff, and researchers around the world. The library collects "collaboratively with peer institutions" and facilitates "international open access, multiplying [its] users' access to materials" (https://library.harvard.edu/about/about-harvard-library). Established in 1638, it is the oldest library system in the United States and the world's largest academic library, employing 700 staff across 28 libraries.