The Alexandria Archive Institute (Open Context)

1. PROJECT JUSTIFICATION

The Alexandria Archive Institute (AAI) seeks an IMLS National Leadership Grants for Libraries – Applied Research Grant in the amount of $699,475 for the project *Advancing FAIR+CARE Practices in Cultural Heritage*. The FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles (Wilkinson et al. 2016) codify fundamental requirements for data intended to serve public needs in science and government. While the FAIR principles focus on eliminating barriers to open data, the CARE (Collective benefit, Authority to control, Responsibility, and Ethics) Principles for Indigenous Data Governance (Carroll et al. 2020) emphasize the need to respect complex social and cultural needs and expectations around data documenting the histories, landscapes, and cultures of Indigenous and other descendant communities. This project will research and develop data management practices for the cultural heritage sector to navigate these two very different visions for data. In doing so, this project addresses program **Goal 5, Objective 5.1** to strengthen the ability of libraries, archives, and museums to work collaboratively for the benefit of the communities they serve by supporting the development of replicable systems that leverage institutional expertise and experience to maximize public access to and use of knowledge resources.

This project brings together participants representing a wide breadth of sectors engaged with cultural heritage data. This collaboration involves leading organizations for archaeological data management, including the applicant (AAI/Open Context), Digital Antiquity (tDAR), the Digital Archaeological Archive of Comparative Slavery (DAACS), the Archaeology Data Service (ADS), and the multi-state Digital Index of North American Archaeology (DINAA) project. Additional constituencies represented include a co-developer of the CARE principles, leaders of the National Association of Tribal Historic Preservation Officers (NATHPO), and others (see *Supporting Document 1- List of Participants*). All will work together to promote the widespread adoption of FAIR+CARE practices in archaeology and cultural heritage among diverse practitioners, including cultural resource management (CRM) firms, state historic preservation offices (SHPOs), Tribal historic preservation officers (THPOs), professional organizations, museums, publishers, governmental agencies, libraries, and data repositories. These new partnerships build upon nearly two decades of the AAI's ethical engagement in cultural heritage data sovereignty issues (Kansa et al 2005; Kansa 2009; see also IPINCH 2008-2016[1]). **This is the first time these diverse constituencies will collaboratively research ways to advance FAIR+CARE practices in this sector. By cultivating these relationships, this project will encourage sustained partnerships beyond the grant period.**

This project aims to reconcile the apparent social and technical contradictions between CARE (highly contextual, socially embedded) and FAIR (fungible, open) principles. By providing clear ethical good practice guidance and digital data governance models integrating FAIR+CARE practices, this project will improve the overall quantity and quality of reusable cultural heritage data, reduce risks of harm, and encourage meaningful participation and benefits-sharing with Indigenous nations and other descendant communities. To do this, the project will establish a network that comprises cultural heritage professionals and policy makers from diverse backgrounds to investigate, define, demonstrate, and promote FAIR+CARE practice across the cultural heritage sector. **The network will explore three research questions:**

Q1. How can we embed FAIR+CARE practices in institutional settings, including libraries, museums, digital repositories, and publishers, as well as federal, state, and Tribal agencies that manage cultural heritage?

Q2. How can data management practices better align interests across different sectors of cultural heritage. Specifically, how can FAIR be adapted to incorporate CARE needs, such as representing Indigenous concepts and worldviews (expressed in Indigenous languages, ontologies, and controlled vocabularies) so they meet Indigenous educational, cultural, and other goals?

Q3. What guides, protocols, training, capacity building, and metadata[2] are needed to make the benefits of FAIR+CARE data more equitably distributed while reducing risks of harmful appropriation?

This project will review current policies, leverage survey data, and use case studies to investigate current practices and needs. This research will inform new practical approaches to ethically responsible curation and effective reuse of cultural heritage data. Multiple outreach strategies will promote, expand, and sustain these outcomes. This effort will

[1] See: Intellectual Property Issues in Cultural Heritage

[2] E.g.: Local Contexts publishes a metadata standard to help communicate Indigenous community needs and expectations.

help build consensus and commitment toward promoting and implementing FAIR+CARE principles across multiple organizational, professional, and institutional sectors. Our network of c. 45 participants, which will divide into working groups to address each of these research questions, will serve as the basis for establishing a longer-term collective engaging with FAIR+CARE across cultural heritage sectors. This work aligns with the specific IMLS program objectives of recognizing and accommodating different values, building a collective set of goals through shared decision-making, establishing shared vocabularies and common practices, formalizing workflow processes, establishing guidelines and standards, building broad infrastructures, and creating project technology.

1.1. Background: The Scale and Scope of Data Challenges in the Cultural Heritage Sector

Conservatively, the public invests over $500 million per year to comply with historical and archaeological protection measures required by federal law (Altschul and Patterson 2010). This level of public investment by itself approaches the total *combined* budgets of the IMLS (roughly $240 million in 2015), the NEH (roughly $200 million in 2022), and the NEA (roughly $160 million in 2022). Since the mid 1960's, cultural resource management (CRM) activities have recorded more than a million archaeological sites, conducted more than one million field studies, excavated over 100,000 sites, and curated more than one billion artifacts and associated records (Ortman and Altschul 2023).

These staggering numbers demonstrate archaeology's importance in public cultural heritage investments. Sadly, much of this work and investment goes largely unnoticed. CRM largely takes place within relatively opaque bureaucratic processes that regulate construction and development. CRM work has resulted in an estimated 350,000 reports nationwide as of 2004 (NADB 2016), but because of poor access and cataloging, irreplaceable cultural heritage documentation in these "gray literature" reports goes unappreciated or is ignored. Recent major federal investments in American infrastructure (~$1.2 trillion for roads, bridges, water systems, renewable energy) will greatly expand the scale and pace of CRM data creation. **Our project to align and improve data management practices in US cultural heritage conservation, thus, has great urgency**.

1.1.1. Data Governance in Cultural Heritage and Other Domains

Cultural heritage conservation spans many sectors, involving Tribal, state, and national government agencies, commercial CRM firms and community organizations, as well as archives, museums, libraries and digital repositories. Archaeological data also have complex technical, professional, social, economic, cultural, legal and policy entanglements. Poor understanding and uncertainty about these entanglements elevate risks associated with data sharing and open science practices. Culturally inappropriate curation practices, as well as data appropriation and misuse can lead to damaging breaches of trust with stakeholder communities and bring harm to these communities.

Many professional domains grapple with sensitive information. The medical and health sciences, genetics, conservation science, psychology, and sociology often create, analyze, and curate datasets requiring anonymization and other controls. Regulatory guidelines, institutional policies, and oversight processes, especially Institutional Review Boards (IRBs), provide invaluable frameworks to guide researchers managing sensitive information in these domains but typically fall short by focusing only on minimizing harms (Saunkeah et al, 2021). IRBs often do too little to promote equity in sharing benefits from participation in research, which in turn, can reinforce existing health, social and economic inequities. Tribal research review boards provide research design toward equitable tangible benefits for tribal communities, protection for both the individual and collective, and their rights to research data that represent them (Gupta et al. 2022). Cultural heritage domains, including archaeology, currently lack professional guidelines for data sovereignty and data governance concerns. In the US, data security attention has typically focused on the geographic location of archaeological resources, particularly archaeological sites and rock-art features. Federal laws, state laws, and professional norms have guarded against public disclosure of such location data under the theory that disclosure would increase risks of vandalism and looting.

Yet, cultural heritage data sensitivity concerns go far beyond location information. Archaeological and museum practice in many parts of the world, particularly in North America, is shaped by long histories of displacement, racially motivated violence, slavery, and colonialism. Power inequalities and abusive practices in the collection, description, and storage of objects and human remains fueled historical divisions and distrust between archaeologists and members of descendant communities (Atalay 2006; Gupta et al. 2020; Murray 2011; Stats 2020). Attempts to

address these past abuses led to national legislation, especially the Native American Graves Protection and Repatriation Act (NAGPRA), that have fundamentally reshaped cultural heritage practices in the US. These examples highlight how archaeology intersects with the legal system, and interests and needs of culturally and historically diverse descendant communities. These intersections also exist in the digital realm, and therefore, archaeology and related fields need to adopt culturally inclusive data governance practices.

Data governance is broadly defined as the "system of decision rights and responsibilities that describe who can take what actions with what data, when, under what circumstances, and using what methods" (Smith, Cruse, and Michener 2011:2). The data governance framework includes knowledge making and strategies for data management; preservation and curation; accessibility; quality issues; and legal and policy concerns over data ownership and data security. The need to protect national security related information, privacy and the rights of human subjects, and the occurrence locations of certain endangered species and historical sites is widely acknowledged across multiple disciplines. Professionals need to also recognize that cultural and historical factors shape how data reflect interests and priorities, and perceptions about sensitivities and risks. To manage culturally defined notions of risk and benefit, data governance practices need to become culturally inclusive (Gupta et al. 2022). Programs to promote data sharing and reuse in archaeology, especially in North America, will fail if they ignore these elements of ethical and social context. **For data reuse and sharing to take hold as a regular and expected mode of professional conduct, data sharing needs ethically appropriate forms of governance inclusive of culturally diverse stakeholders.**

The ethical treatment of archaeological and other cultural heritage data is a topic of frequent discussion (Kansa 2009, 2016; Kansa et al. 2005; Kansa et al. 2018; Kintigh 2006; McManamon and Kintigh 2010) that has resulted in new policies and revised ethical statements by professional or trade organizations such as the American Cultural Resources Association (ACRA 2019) and a Register of Professional Archaeologists (RPA) effort in progress. Moreover, as archaeological field research has become more expensive and subject to political and other restrictions, there is an increasing demand for reuse of existing data to inform research, instruction, and conservation policies.

Coupled with these internal pressures for data reuse are closely-related external pressures to increase accountability and transparency in research, notably in open access (e.g., The White House 2013; G8 Open Data Charter (2013); OSTP 2014; OMB 2015; the Open Government Data Act, PL 115-453, Title II); research replicability though journal requirements for provision of supporting data (Nature 2013; Vines et al. 2013); and effective data management that leads to effective data discovery, access, reuse, and long term preservation (e.g., NSF and NEH requirements for Data Management Plans). These efforts align with long-standing ethical principles in archaeology and other heritage professions concerning public ownership, stewardship, public education, and conservation of cultural resources.

Yet, despite more than ten years of availability of reputable repositories and data publishers (e.g., ADS, Open Context, and tDAR), numerous publications arguing for their use (e.g. Richards 2017; Kansa 2012; Kansa et al. 2013; Kansa et al. 2018; Marwick and Birch 2018; Kintigh et al. 2015; McManamon and Kintigh 2010; McManamon et al. 2017), and nearly a decade of funder-required data management, only a miniscule fraction of recent, publicly funded (or compliance mandated) field projects have data that are findable, accessible or reusable. Digital data collected at enormous expense are continually, rapidly, and permanently being lost due to hardware failures, media degradation, software obsolescence, and inadequate documentation. Much of what is not yet lost, is not discoverable or accessible and will ultimately be lost unless data governance practices change dramatically (Kintigh 2006; Kansa et al. 2014; Michener et al. 1997; Richards and Moore 2015; Vines et al. 2014; Faniel and Yakel 2017). This problem is not new, but tragically, it is not noticeably improving—to the detriment of research, public education, and cultural heritage conservation. **This project seeks to dramatically change this discouraging trajectory by developing protocols, data governance processes, and good practice guidance to ethically and responsibly implement FAIR+CARE data governance principles in cultural heritage.**

1.1.2. Why FAIR+CARE?

CARE principles center people and purpose in data governance whereas FAIR principles are data-centric. To facilitate implementation of both FAIR+CARE principles, we start from the outset of research design, which can include conversation about human and machine-readable cultural metadata for digital archaeological data (Gupta et al. 2023). This represents a key change in how practitioners collect and document archaeology, and will have

immediate outcomes in building well-documented, curated data that maintain provenance, or connection with communities. Provenance will facilitate the application of consultation processes, possible access controls, and communication of other expectations based on cultural protocols throughout the data lifecycle (Carroll et al. 2021).

The FAIR Guiding Principles for scientific data management and stewardship were first published in 2016 with the goal of improving the infrastructure for data reuse (Wilkinson et al. 2016). The principles represent guidelines for the management of public data that, if widely adopted, would transform the impact of data sharing by improving data Findability, Accessibility, Interoperability, and Reuse. European cost-action efforts, such as the Saving European Archaeology from a Digital Dark Age ARIADNEplus data portal (Richards et al. 2021, 2022), demonstrate well-funded, large-scale, international efforts by governments and organizations to make data from archaeological investigation FAIR. Outside of Europe, there are several exemplary individual efforts to make specific datasets, information, and even computational code FAIR in archaeology (Davies et al. 2021; Fritsch 2021; Hiebel et al. 2021; Marwick and Wang 2019; Nuninger et al. 2020; Schmidt and Marwick 2020); however, broader, discipline-wide efforts to explicitly address the FAIR principles and practices as they pertain to datasets, policy, and protocols in the Americas are still largely absent. Furthermore, in the absence of CARE Principles, the implementation of FAIR in cultural heritage will be problematic and is unlikely to have wide adoption.

The CARE Principles for Indigenous Data Governance, crafted by the Global Indigenous Data Alliance (GIDA), offer a critically needed ethical framework for stakeholder centric governance in the stewardship of culturally and geographically sensitive data (Carroll et al. 2020; Carroll et al. 2021; Gupta et al. 2020). Making space for cultural metadata, and digital tagging through Traditional Knowledge (TK) labels, Carroll et al. (2020) remark that "the development process for CARE should necessarily align with FAIR given the various ways in which the principles might intersect around different datasets or contexts." Cultural metadata might include three components: provenance, protocol, and permissions (Anderson and Hudson 2021). Good practices in the use and reuse of globally unique persistent identifiers play an important role in facilitating interoperability and reuse. Such identifiers are used to document data with shared concepts and standards, and they enable collection of data citation and impact metrics. Globally unique persistent identifiers can also address CARE principles by expressing and communicating aspects of contextual integrity. Such identifiers can play a key role in maintaining provenance and associations between specific data resources and community-defined protocols and expectations for their use (Kansa and Kansa 2022).

This project highlights the synergistic complementarity of FAIR+CARE data practices which can be used to advance the understanding and conservation of cultural heritage. Cultural heritage professionals face uncertainty about how to ethically understand and respond to data sensitives shaped by complex historical, economic, cultural, and political factors. This uncertainty undermines efforts to promote FAIR practices. The Registry of Research Data Repositories, hosted by DataCite lists 70 digital repositories, globally, related to "Ancient Cultures", whereas life sciences have 1,526 and natural sciences 1,390. The limited number of datasets housed in these repositories alone indicates that FAIR data principles are not widely embraced in the cultural heritage sector. Guidance for how to practice FAIR+CARE in tandem, together with clear data curation governance processes, will help reduce the risks and uncertainties in data management. Practices that promote cross-cultural collaborations, shared understanding about data sensitivities, and building shared consensus in how to use data for research, community engagement, and conservation outcomes will all support more ethically appropriate and equitable participation in data sharing.

2. Project Work Plan

Application of FAIR+CARE principles in the cultural heritage sector will require purposeful effort, greater levels of transparency and collaboration, and especially, more meaningful and sustained partnerships with Indigenous and descendant communities. This proposal supports a **FAIR+CARE Cultural Heritage Network** of working groups and outreach efforts that will result in a detailed specification of the implications of FAIR+CARE practices for different types of cultural heritage organizations and communities; development of implementation strategies with them; and a plan to promote a widespread commitment to the FAIR+CARE principles by cultural heritage organizations, governmental agencies, cultural resources management firms, museums, digital repositories, publishers, and funders. [See *Schedule of Completion*]

2.1. Research Methods

This project will use complementary research methodologies involving a broad survey and specific case studies to gather information about FAIR+CARE data management issues.

2.1.1. Research Method 1: Online Survey of Cultural Heritage Professionals

To help build a wide breadth of understanding, we will develop an **anonymous online survey to gather qualitative and quantitative information about incentives, opportunities, and challenges from respondents in the cultural heritage sector**. The survey will be developed (with IRB approval) and broadly shared by all participants among their own communities and networks. The results of the survey will inform priorities for more in depth case study investigations. The survey will explore the following topics (see example questions in the *Digital Products Plan*):

- **Data Ownership**. In heritage management contexts archaeological data and work products (e.g., technical reports) are often subject to sharing conditions and limitations that are governed by the contracting agency, proponent, and/or permitting agency. Cultural resource databases created by Indigenous agencies face similar concerns. How do legal concerns impact data exchanges between an organization and external partners?
- **Access and Control for Sensitive Data** is needed where demanded by law or professional ethics. The CARE principles take into account concerns of descendant communities over access to sensitive information, and loss of connection to collective authority over data, an issue highlighted by the Global Indigenous Data Alliance (2020). How do organizations currently respond to this need?
- **Interoperability/Data Integration Issues**. The lack of data standards and use of shared, controlled vocabularies by data creators presents major challenges in data discovery for repositories and for data integration by humans and machines (Kintigh et al. 2018). What barriers or challenges exist in the meaningful use of well-established standards (e.g. Dublin Core metadata, CIDOC-CRM, etc.) exist in research infrastructures, including data federators, digital repositories, and their user communities?
- **Proprietary Database Software** can present challenges for both interoperability and conversion to long-term data preservation formats. What software do cultural heritage institutions use and how well do these tools support interoperability and open, nonproprietary formats?
- **Transactional Databases**, such as those for archaeological science samples present challenges for both practical citation with persistent identifiers and for archiving and long-term data preservation because they are continually changing. How do organizations manage archiving and version control of transactional databases?
- **Recognition**. Standards for attribution of credit (notably, Creative Commons licenses) are well established but present some practical difficulties. How, in practice, is an integrated dataset derived from dozens of source datasets properly cited and credited? Moreover, what metadata, persistent identifiers, and other practices need to be in place to recognize the authority and interests of Indigenous communities and other descendant communities (Montenegro 2019)? More subtly, how is it possible to retain the information regarding data attribution and stakeholder community recognition in linked open data contexts?
- **Contextual Integrity** is often poorly maintained in cultural resource datasets. Heritage management activities often involve "handoffs" of objects (artifacts, ecofacts) between institutions with very different data management systems. What frictions may make provenance more difficult to track and may impede collaborative curation and repatriation work involving Indigenous communities?
- **Workflow Imbalances** (e.g., delaying data archiving to the end of the project when needed knowledge is no longer available or project funds have been exhausted). How can good practices that result in FAIR+CARE data stewardship become firmly embedded in project workflows in both academic and heritage management contexts from the start of the data lifecycle?
- **Training and Data Literacy Inadequacies**. How do we facilitate knowledge of FAIR+CARE data stewardship practices and train students and active practitioners about purposeful and ethical workflows in cultural heritage professional roles?
- **Unfunded mandates:** Who will incur the financial costs associated with making data FAIR+CARE? Will these costs fall to public agencies, an Indigenous community, or grant or client funding?
- **Disparate repositories:** At present, there is no singular digital archive which secures and makes accessible all digital archaeological datasets. In addition to dedicated, non-profit digital archaeological repositories (e.g. tDAR, Open Context, DAACS), all state historic preservation offices have their own versions of a cultural resource database, which often consists of no more than a hard-drive with the raw files and no associated metadata. The

situation is similar for THPOs and academic projects that maintain databases and datasets directly related to, or about descendant communities and Indigenous peoples, their lands, waters, resources and knowledge systems. What barriers exist in using data curation infrastructure outside a given institution?

- **Collective Benefit:** Non-Indigenous use of Indigenous data has led to harm for Indigenous communities. Tribal agencies and Nations must derive collective benefits from FAIR+CARE data practices. What tangible benefits could be gained for Tribal agencies (leadership roles, capacity building, better information for decision making) can come from using (appropriately) open and interoperable cultural resource data?

While it is not feasible for this project to solve all of the above data management issues that may be explored in the survey, the survey will help prioritize concerns where the projects can feasibly make incremental progress on promoting FAIR+CARE practices.

2.1.2. Research Method 2: Case-Study of the Curation of Specific Datasets

In addition to the survey, we will use case studies to examine specific datasets curated by project participants to develop more in depth and qualitative understanding of FAIR + CARE data issues. The case studies will investigate many of the same topics as the survey, but through the lens of deeper interrogation of data curation practices of specific datasets. Expert representatives of various sectors will examine a range of datasets including:

- *Large geographic and chronological scope*: Some cultural heritage datasets such as DINAA document cultural resources that date from the Pleistocene to the present across North America. As such, potentially hundreds of Native American nations as well as other descendant communities (such as Black communities, various religious and diaspora groups, veterans and their descendants, etc.) may have an interest in the governance of DINAA. Large scale datasets can inform governance processes led by groups such as NATHPO, which collectively represents the cultural resource management interests of Native American nations across the US.
- *Thematic data*: DAACS has developed a large-scale database documenting the history and archaeology of enslaved peoples in the Americas. The lasting economic, social and cultural impacts of institutionalized slavery play a significant role in contemporary society. DAACS can provide a valuable case study to inform inclusive approaches to managing data about historical epochs that powerfully impact living communities (Galle et al. 2019), opening conversation on community governance of heritage where formal legal protection falls short.
- *Museum data*: Museums and other (artifact) collections repositories are responsible for the storage and curation of objects recovered in cultural resource management. Museum databases may lack interoperability with archaeological databases documenting their discovery. These frictions make provenance more difficult to track and may impede collaborative curation and repatriation work involving Indigenous communities. tDAR currently archives some 375 museum related datasets that may each provide an informative case study. A detailed case study of museum collections' data lifecycles can help identify better data management and governance processes.

2.1.3. Research Method 3: Analysis of Existing Documentation and Data Management Guidance

The final research method will involve **documentation and analysis of relevant publicly available data management policies and documentation in the cultural heritage sector**. In addition, Working Group 3 ([WG3] see below) will examine existing self-assessment and interactive guidance tools already deployed to guide development of Data Management Plans (DMPs) and FAIR data. WG3 will also examine interactive guides and assessment tools created to guide the management of sensitive information, such as data gathered in studies under IRB oversight. Study of existing public documentation and guidance tools will inform WG3 objectives in developing an interactive self-assessment tool to help guide FAIR+CARE practices.

2.2. Working Groups

The research methodologies discussed above will help identify current deficiencies and approaches to improve FAIR+CARE practices appropriate to local, thematic, and large continental-scale contexts. These studies will inform identification of governance models, metadata documentation (such as LocalContexts labels and notices), outreach and consultation needs, and other issues. This three-year project will organize ~45 project participants into three Working Groups (WGs). Each WG will have its own area of topical focus and will meet separately to both conduct research (each group using one methodology discussed above) and produce deliverable outcomes / work products. To

maintain overall cohesion and the cross-pollination of ideas, WGs will come together quarterly over Zoom and all participants will meet at in-person workshops at the end of Years 1 and 2. Project co-directors and advisors, who will represent each working group, will also meet monthly on Zoom, in addition to the annual all-hands meetings.

Broadly speaking, in Year 1, the project will mainly emphasize information gathering activities using the research methodologies discussed above. Specifically, these include survey development and data collection (WG1), case study selection and assessment (WG2), and study of existing policy documents and DMP (and IRB related) guidance tools (WG3). Year 2 will focus on synthesis and development of new policy recommendations and supporting demos and tools, and Year 3 on refinement, dissemination, and promotion.

2.2.1. Working Group 1: Digital Data Governance Models

Q1: *How can we embed FAIR+CARE practices in institutional settings, including libraries, museums, digital repositories, and publishers, as well as federal, state, and Tribal agencies that manage cultural heritage?*

WG1 will be responsible for developing the survey at the start of the project, in (virtual) collaboration with the co-directors and advisors (see Online Survey of Cultural Heritage Professionals, above). WG1 will synthesize research results to inform **practical and actionable inclusive digital data governance models** for use by organizations and agencies across the cultural heritage sector.

Transient interactions do not meet "protocol" and "permissions" requirements of CARE data. Instead, they require lasting maintenance of ties between institutions and communities (Nicholas and Gupta 2022). Thus, this project aims to foster lasting commitments and relationships among participants and the wider cultural heritage community by providing opportunities to connect and develop new connections moving forward. This WG will also codify insights gained from its research into a **FAIR+CARE Cultural Heritage Data Commitment** and encourage individuals and organizations to sign on. WG1 will recommend establishing an interest group in a professional society to review the Commitment and support this network into the future. The Commitment will be available to read and sign on the FAIR+CARE website. The Network for Publishing Data in the Earth and Space Science's commitment statement is an excellent model for a likely outcome (COPDESS 2020). The body of signatories will establish a basis for establishing interest groups within the SAA, SHA, and/or NATHPO to transition the project to a self-sustaining network and ensure updates and continued promotion of these efforts.

Open Context, tDAR, DINAA, and DAACS will integrate these governance models into their data curation practices and report lessons learned in the process to the wider community through publications and public presentations. Adoption of FAIR+CARE governance practices across these organizations will institutionalize and help sustain this project's outcomes. For example, the related *Intellectual Property Issues in Cultural Heritage* project (2008-2016) informed development of Open Context's intellectual property policies – a lasting impact years after it concluded.

2.2.2. Working Group 2: Promoting and Demonstrating Collaborative Practices

Q2*: How can data management practices better align interests across different sectors of cultural heritage. Specifically, how can FAIR be adapted to incorporate CARE needs, such as representing Indigenous concepts and worldviews (expressed in Indigenous languages, ontologies, and controlled vocabularies) so they meet Indigenous educational, cultural, and other goals?*

WG2 will have the primary responsibility of developing in-depth case study methodologies discussed above. Informed by the survey results and synthesis of case study research, WG2 will pilot and disseminate **collaborative practices to support Indigenous community values and interests** in appropriately representing Indigenous languages and concepts in ontologies, controlled vocabularies, and metadata in cultural heritage data, as well as public engagement and teaching materials presenting such data.

WG2 will produce guidelines and training documents, including examples of FAIR+CARE in practice, as well as sample data management plans that incorporate FAIR+CARE practices. This work will leverage and enhance closely related programs such as the Digital Stewardship Curriculum, the Trust Principles for Data Repositories (Lin et al. 2020), and the Repatriation Meets Protocols Workbook and Resource Guide.

Examples of FAIR+CARE in practice will involve building on existing public data publishing, curation, and analysis services to implement Local Contexts metadata and traditional knowledge (TK) labels across the various systems (tDAR, Open Context, DINAA, DAACS, etc.). To maximize interoperability with existing infrastructure, we will define common conventions that use widely supported Dublin Core Terms attributes (esp. *provenance*, *rights*, *rightsholder*, *license*, and *source*) to express CARE related metadata aligned to Local Contexts recommendations. Participating collections in this network will coordinate their implementations to promote interoperability, then share their implementation strategies with other networked scientific and digital library data infrastructure (via other partnerships outside the scope of this specific proposal with systems such as ZooArchNet, Neotoma, and iSamples).

To facilitate uptake and use of FAIR+CARE practices, a demonstration implementation of **FAIR+CARE promoting data models and metadata will be developed and demonstrated using Arches**. Arches is an open source cultural heritage data management application used for both cultural resource inventory (historical structures, archaeological sites, landscapes, etc.) and museum collections (especially object conservation) [see *Digital Products Plan*]. Arches has great flexibility for modeling data and defining controlled vocabularies, including culturally and linguistically diverse concepts. The working group also will identify relevant datasets curated by participating systems that would demonstrate both data integration related scientific research outcomes and inclusive practices for data documentation, data governance, and benefits sharing. Guidelines, links to examples, and training materials will be delivered on a dedicated FAIR+CARE website, promoted via social media, listservs and professional societies, and reviewed and updated on a regular basis by dedicated interest groups in professional societies participating in this network.

2.2.3. Working Group 3: FAIR+CARE Cultural Heritage Data Assessment Tool

Q3*: What guides, protocols, training, capacity building, and metadata are needed to make the benefits of FAIR data more equitably distributed while minimizing risks of harmful appropriation?*

WG3 will begin with a review of existing public policy documents, FAIR data tools, data management plan tools, and any interactive guidance tools related to CARE (such as sensitive data "checklists" and assessment tools). Informed by survey results and synthesis of case study research, in Years 2 and 3, WG3 will design and refine guidelines and training materials packaged into a **FAIR+CARE Cultural Heritage Data Assessment Tool** (a simple GitHub hosted, client-side Web questionnaire built with Bootstrap and Vue.js) that will suggest ethical data practices customized to user responses. This will support practitioners in self-assessment of their alignment with FAIR+CARE practices and suggest possible strategies to achieve better alignment.

The FAIR+CARE Cultural Heritage Data Assessment Tool that will guide users in identifying good data practices and opportunities for improvement in their digital datasets. While FAIR assessment tools exist (e.g. the Dutch Data Archiving and Networked Services (DANS) and the Australian Research Data Commons (ARCD), among others, have developed FAIR data self-assessment tools), none to date incorporates CARE practices. WG3 will build on existing FAIR assessment tools to develop an assessment tool for the cultural heritage sector that incorporates considerations of CARE practices. Through a series of brief questions, a user may qualitatively assess "How FAIR+CARE" their data are. Their resulting score will include a summary of areas that need improvement, as well as potential steps they may take to incrementally improve the FAIR+CAREness of their data. This process will provide people with implementable solutions that lead to ethically responsible curation and effective reuse of archaeological and other cultural resource data. The tool will be hosted on the FAIR+CARE website (developed by the project) and will be promoted through the network, social media, public presentations, and publications in such venues as *SAA Archaeological Record*, the American Library Association's *American Libraries Direct* newsletter, and the American Alliance of Museums *Museum Magazine*. The tool itself will be released open source (GPLv3 license) and will be under public version control with GitHub.

2. 3. FAIR+CARE Cultural Heritage Network: Structure and Interactions

The full **FAIR+CARE Cultural Heritage Network** established by this project will be a collaborative structure to support the communication and coordination of FAIR+CARE-related practices across disciplinary, organizational, divisional, and geographic boundaries. The network brings a balance of synergistic strengths from libraries, cultural resource management firms, data repositories and publishers, museums, agency and regulatory representatives,

professional societies, academic organizations or projects curating and reusing synthesized data, educators, and Tribal nations. The Directors and Advisors bring diverse expertise from Tribal (including leadership of NATHPO), Indigenous data governance (including the co-lead of the CARE principles), CRM, public policy, and data repositories. The network is structured as follows (see *List of Key Staff and Consultants* and *Supporting Document 1-List of Participants* for participant names and affiliations). The budget covers travel for all participants and stipends for their participation that reflect the anticipated time commitment per category.

- ***Project Directors:*** Four **Project Directors** will be responsible for oversight of the project, including coordination of the full network and WGs, and for ensuring that project deliverables are met. They will meet virtually on a monthly basis with the Advisors, on a quarterly basis with WGs, and more frequently with each other, as needed.
- ***Project Advisors:*** Collaborating with the Project Directors in quarterly virtual meetings, eight **Project Advisors** will contribute their expertise and coordinate / oversee the three WGs, and manage the WG outcomes.
- ***Network Participants:*** Approx. 35 additional **Network Participants** will select one WG to join for the duration of the project. 12 people per working group ensures a balanced distribution of expertise and sectors represented across WGs while also providing a buffer in the event of turnover. Network Participants will participate in quarterly virtual meetings and two in person workshops over the course of the project.

The full network will gather at two workshops at the end of project Years 1 and 2. These in-person meetings will provide an opportunity to discuss and cross-fertilize WG insights across the full project network. We have achieved diversity with our participant list, and the in-person workshops will promote equity and inclusion by prioritizing time for discussions rather than lectures. Each workshop will involve an all-hands meeting followed by a series of breakout meetings for the three WGs. We will follow a "scrum meeting" model where the WGs answer three questions to the full group: (1) What did we do yesterday to move our project forward? (2) What are we doing today to move our project forward? (3) What is standing in the way of our project moving forward? (after Watrall 2019). This provides iterative group feedback to ensure that goals are set and progress is made at these key in person meetings. Each WG will have quarterly virtual meetings between in-person meetings, where WGs will be able to bring in additional participants should they seek additional expertise or input.

3. DIVERSITY PLAN

The composition of our leadership team has informed proposal planning. Three of the four PIs are women, two of the four are racialized, one is an Indigenous CRM professional, and one is an expert in the application of CARE in archaeology in support of Indigenous knowledge and data sovereignty. Our full network of participants represent numerous constituencies (some people with multiple): academic settings (15), Tribal perspectives (9), data repositories (5), libraries/research support (5), museums (4), public agencies (9), the cultural resource management industry (6), professional organizations (6), and publishers (1). Twenty participants are women and twelve are from the BIPOC community. The postdoctoral scholar is a racialized woman. Participant organizations represented include the Society for American Archaeology, Society for Historical Archaeology, UC Berkeley libraries, C.E. Smith Museum at Cal State East Bay, Badè Museum of Archaeology, Register of Professional Archaeologists, National Park Service, Bureau of Land Management, Army Corps of Engineers, National Association of Tribal Historic Preservation Officers, and State Historic Preservation Offices, among others.

The project's budget reflects its commitment to diversity and inclusion. Uncompensated labor, especially labor contributed from representatives of marginalized communities, would undermine the ethical goals of this project. Instead, the budget explicitly compensates participants for their commitments of labor and expertise. Support for network participants with stipends and travel makes up the majority (over 65%) of the budget.

Participants represent both early career and established professionals, providing diverse perspectives for the full network and to populate each of the three WGs. Our meetings will be both in-person and virtual to accommodate the work and travel needs of our diverse participants. **This is the first time that representatives of such diverse cultural heritage stakeholders (academic, museums, libraries, CRM, public, Tribal) have come together to shape data management practices and policies.** The four project co-directors and eight advisors have a history of working to address the challenges and opportunities of computational infrastructure in support of cultural heritage and working with Indigenous communities on data sovereignty. These individuals will divide evenly among the three WGs to provide guidance and expertise, as well as ensure cross-pollination and coordination of outcomes of all WGs

in regular virtual meetings over the course of the project. By leveraging our network to share the survey widely across their communities, we will ensure that we collect a robust dataset of diverse perspectives to inform our collective outcomes. **This unprecedented collaboration across many sectors and voices establishes a new model for inclusive and equitable practices in the management of cultural heritage data.**

4. Project Results

This project will develop, disseminate, and promote ethical good practice guidance and digital data governance models integrating FAIR+CARE practices. Public presentations, research papers in refereed journals, articles in professional newsletters, a public website with clear and practical guidance, a reference implementation (demo with Arches), and an online self-assessment tool and guide will disseminate ways to improve the overall quantity and quality of reusable cultural heritage data while reducing risks of harm. Project participants from a wide breadth of organizations, professional sectors, and communities will leverage their far-reaching personal and professional ties to encourage meaningful participation and benefits-sharing with Indigenous nations and other descendant communities in data management. Moreover, these sustained multi-stakeholder collaborations will catalyze continued partnerships to **identify priorities for future investments to expand upon this work beyond this grant period.**

By integrating FAIR+CARE, this project will demonstrate that excellence in data curation is not, and cannot be, a program to alienate digital data from stakeholders solely for the benefit of narrow technocratic interests. Open Science and Open Government practices that promote excellence and transparency can and should be adapted to support continued inclusion of stakeholder perspectives and interests in cultural heritage data. By demonstrating how FAIR+CARE practices work synergistically, this project will highlight how data curation and reuse is not only relevant but central to the ethical conduct of cultural heritage conservation, teaching, and research today.

Immediate outcomes of this project include the establishment of a diverse network of individuals and institutions to promote FAIR+CARE practices, as well as training materials, an assessment tool, and an open-source reference implementation (via Arches) to demonstrate and facilitate adoption. However, this project will also catalyze more lasting and sustained improvement in data curation practices. By incubating a longer-term coalition across key institutions, future partnerships will oversee updates to these models and tools as needs evolve. Longer term impacts of this work include the following:

- ***Enhanced Infrastructure Partnerships.*** This project cultivates partnerships across key cultural heritage information systems and extends these partnerships to academic, museum, library, Tribal, and public sectors.
- ***Socially Responsive Data Governance and Management.*** This project invites sustained, purposeful participation and leadership from members of communities often sidelined from policymaking and science. Cultural heritage data not only has scientific and educational value; it also informs policy decisions on the conservation and protection of resources threatened by development, climate change, vandalism, and neglect. Our approach will support science in becoming more responsive to the needs and interests of a broader and diverse range of social groups. Ultimately this will lead to less adversarial and polarizing conduct of science and policy making, especially in cultural resource conservation.
- ***Technical.*** A key outcome of the project working groups is the development of publicly accessible, online resources of technical guidance for making data FAIR+CARE. A demonstration implementation using the open-source Arches cultural heritage data management application will facilitate wider adoption.
- ***Education.*** The project will outline the educational needs and steps necessary to improve data stewardship across the cultural heritage sector. Its website on FAIR+CARE data practices will help disseminate critically needed training materials and demonstration implementations. These materials will encourage professional development and help build capacity to institutionalize greater professionalism, quality, and ethical conduct in cultural heritage.
- ***Increased Public-Private Partnerships***. Most cultural heritage data derive from research conducted by private-sector firms in response to governmental mandates for mitigation of impacts to cultural resources being damaged or destroyed by development. Adoption of FAIR+CARE practices, as well as expanding the conversation to include diverse voices across academia, memory institutions (libraries, museums, and archives), and the public and private sectors, will lead to greater impact and uptake of this project's outcomes.

The Alexandria Archive Institute (Open Context)

Schedule of Completion

Year 1: 8/1/23 – 7/31/24

Year 1 Activities	Q1	Q2	Q3	Q4
All hands virtual kick-off meeting. Network members choose WGs (first half of Q1)				
Post-workshop evaluation meeting (co-directors, advisors) (second half of Q1)				
Planning & Logistics for in person workshops (co-directors and advisors)				
All hands in-person workshops (full network)				
Project website (AAI staff, consultants) Setup & updates				
WG1: Digital Data Governance Survey: Design, review, and circulate before first all hands meeting (quarterly virtual meetings)				
WG2: Collaborative Practices Dataset Case Studies: Select and assess case studies from network members; prepare for sharing at workshops (quarterly virtual meetings)				
WG3: FAIR+CARE Data Assessment Tool Study existing policy documents and guidance tools (quarterly virtual meetings)				

Year 2: 8/1/24 – 7/31/25

Year 2 Activities	**Q1**	**Q2**	**Q3**	**Q4**
Planning & Logistics for in person workshops (co-directors and advisors)				
All hands in-person workshops (full network)				
Post-workshop evaluation meeting (co-directors and advisors)				
WG1: Digital Data Governance Synthesis & development of policy recommendations; drafting of commitment				
WG2: Collaborative Practices Development of demos, training materials				
WG3: FAIR+CARE Data Assessment Tool Development of tools				
Project website (AAI staff, consultants) Setup & updates				

Year 3: 8/1/25 – 7/31/26

Year 3 Activities	**Q1**	**Q2**	**Q3**	**Q4**
Post-workshop evaluation meeting (co-directors and advisors)				
WG1: Digital Data Governance Refinement of recommendations & commitment				
WG2: Collaborative Practices Refinement of examples, training materials				
WG3: FAIR+CARE Data Assessment Tool Refinement of tool				
In person wrap up meeting (co-directors and advisors)				
Dissemination of outcomes – materials, publications, public communications, assessment tool, commitment (co-directors and advisors)				
Project website updates, FAIR+CARE Data Assessment Tool posting (AAI staff, consultants)				
Establish interest group at a professional society (co-directors and advisors)				

The Alexandria Archive Institute (Open Context)

Digital Products Plan

This project will result in several types of digital products (availability, access, and sustainability are addressed for each of the following products):

1. Reports and public presentations: Project participants will develop a number of presentations and documents. Some of these will report on the progress of each Working Group during the project. Some presentations and reports will be developed for communication and promotional purposes. In general, presentations and reports will be made available open access under a Creative Commons Attribution License. However, if required by stakeholders, some content may be distributed under more restrictive licensing conditions. Zenodo.org and tDAR will provide long-term preservation and access of such content as well as DOI identifiers to facilitate citation.
2. Online Survey of Cultural Heritage Professionals: The project will develop a Qualtrics hosted online survey instrument available through the UC Berkeley licensing agreement with Qualtrics. Qualtrics provides data validation and "skip logic" features that show or hide different survey questions depending on responses made earlier. The survey will be developed and circulated to cultural heritage professionals, especially those with interests and roles in data management. The project team will submit the survey for Institutional Review Board (IRB) review and ensure that respondent privacy and other risks are addressed. To maintain respondent privacy, survey results will be used internally by members of Working Group 1. IRB approved aggregate data (with low risks for personal identification) will be made available through Zenodo.org in one or more CSV files. Draft (subject to revision) survey questions include the following:
    a. Size and capacity of the organization
    b. What kinds of interactions do you have with heritage data?
    c. What software or digital tools do you currently use to manage, process and share data?
    d. (How) do you currently share data with others within your organization and outside your organization?
    e. (How) do you manage, use, and share data that relate to Indigenous and other descendant communities?
    f. What is your familiarity with FAIR?
    g. What is your familiarity with CARE?
    h. What sensitivities are you aware of in the data that you manage, and what steps do you take to minimize misuse of data, and maximize benefits from data?
    i. What barriers or difficulties have you faced in accessing and reusing data from an external source? / or sharing data deposited by an external client?
    j. Are data you have reused or shared well documented with information on its provenance (where data come from; which community to engage with regarding consent for use and future use)?
    k. How do you document permissions and protocol information (guidance on appropriate access, use, and care of data, and potential harm and benefits associated with data based on Indigenous community's worldview)?
    l. How confident do you feel in the quality of the relationships you have with stakeholders related to the data that you manage?
    m. How do stakeholders and internal and external partners you engage with use the data you manage? How easily do you exchange data with partners? If not, why?
    n. Could your data be of more value to external organizations, individuals, and/or communities?
    o. How confident are you in the long-term preservation and access of the content you manage?
    p. What issues would you like to understand better to make the information you manage more valuable to external partners and stakeholder communities?
    q. What incentives do you have for making your data accessible outside your institutional setting? What are the barriers in making it accessible?

3. FAIR+CARE Website: This project will develop a simple (static HTML, CSS, some Javascript) website hosted with GitHub. We will use GitHub to make version control and issue tracking (including community feedback) convenient. Zenodo.org also easily integrates with GitHub, facilitating long term preservation and archiving the project website. To reduce web development costs, the project website will use Bootstrap (or similar CSS + Javascript) frameworks for responsive, mobile-friendly user experience. The website will link to presentation and document files on Zenodo, the *FAIR+CARE Cultural Heritage Data Commitment* (see below) and the *FAIR+CARE Cultural Heritage Data Assessment Tool*. The domain name will be owned by the Alexandria Archive Institute (Open Context) or transferred to another suitable nonprofit organization for at least 10 years. The management of specific parts of the project website include:
    a. FAIR+CARE Cultural Heritage Data Assessment Tool: The *FAIR+CARE Cultural Heritage Data Assessment Tool* will be a client-side (Javascript) only interactive Web based form that will provide data management guidance tailored to user responses to input forms. The open source Bootstrap + Vue.js will be used to create responsive, accessible and mobile-friendly forms and responses. Along with the rest of the website, this tool will be under Git version control and all code will be made available under an open source (GPLv3 license) license. The form will not collect or persistently store any user provided information. All user response data will be maintained temporarily on the user's own browser.
    b. FAIR+CARE Cultural Heritage Data Commitment: The *FAIR+CARE Cultural Heritage Data Commitment* will be a short HTML document (roughly 1 page printed) that, drawing upon project research outcomes, defines key principles and practices in cultural heritage data management that individuals and organizations can review and publicly endorse. Version control and issue tracking will facilitate drafting and future revisions of the document.
    c. FAIR+CARE Cultural Heritage Data Commitment Endorsements: Endorsements of the *FAIR+CARE Cultural Heritage Data Commitment* will be represented in a simple Comma Separated Values (CSV) table listing names and affiliations of individuals and organizations endorsing project developed FAIR+CARE data management policies. The CSV data will be maintained publicly on GitHub. Members of the community can add their endorsement through either a Google Form or a "pull request" via GitHub. Open Context administrators will be responsible for periodically (initially weekly then on a less frequent basis) adding entries collected with Google Forms to the GitHub hosted "canonical" dataset. The dataset will be archived with Zenodo.org where versions will be updated and tracked.
    d. Web Analytics (Matomo): To capture Web user analysis, we will use the open source Matomo analytics suite. Matomo functions much like Google Analytics, but offers better user privacy protections. We also use the recommended privacy default settings for Matomo to limit the specificity of data collected about users. Limiting the specificity of data collection reduces confidentiality risks to individuals in the event of an accidental or malicious breach of user data.Our goal is to limit risks associated with data collection and will only collect and retain the low-sensitivity data described above. These data will be used to help assess impact and help guide improvements to the FAIR+CARE Website.
4. Arches FAIR+CARE Demonstration Site: Arches (https://www.archesproject.org/) is an open-source Web application built with Django (Python) framework. Designed to meet data management needs in the cultural heritage sector, Arches supports sophisticated and flexible data modeling. It is widely deployed by public agencies with regulatory oversight of historical resources, by research teams, and by museum curator and conservators. Project use of Arches will involve several aspects, each with their own planning needs:
    a. Arches powered website (https://arches.opencontext.org): The project will develop a small cloud computing hosted Arches deployment to demonstrate a FAIR+CARE-aligned structured dataset. Open Context will maintain the Arches deployment for at least five years.
    b. CARE metadata graph: To support Linked Data and facilitate implementation of ontologies (especially the CIDOC-CRM) Arches models data in graph structures. Arches users can define and share "branches" that provide a common structure for attribute information. The project will define a branch to express elements of metadata that support CARE needs. This branch will be demonstrated on the arches.opencontext.org instance. The branch will also be available to other Arches users as a reusable open source JSON expressed configuration disseminated (and curated) via the Arches Resource Model Working Group (https://www.archesproject.org/arm-wg/). Other Arches implementers will therefore be able to install the CARE metadata branch to model data on their own Arches powered systems.

The Alexandria Archive Institute (Open Context)

Data Management Plan

This project emphasizes metadata documentation of existing datasets and the development of policy guidelines and instructional materials. Regarding data management, this project will:

1. research, develop, and document ways to enhance CARE-related metadata documentation of existing and future archaeological datasets;
2. develop and disseminate policy and governance guidelines as well as professional development and instructional materials; and
3. develop and deploy a Web-based FAIR+CARE Cultural Heritage Data Assessment Tool.

Structured Data and Metadata

This project will not create, manage, or serve new (structured) archaeological research datasets. The project will leverage existing data already curated by tDAR, Open Context, and DAACS. Data in these platforms already have well-established data curation and preservation processes. tDAR itself is a digital repository (with options for open and restricted access) and co-developer along with the Archaeology Data Service (ADS) of the discipline's leading data documentation standards. Open Context, an open access data publisher, documents and annotates datasets using Linked Open Data methods, and archives standards-documented data (in open JSON-LD and CSV formats) with the University of California's Merritt Repository and Zenodo. Open Context also hosts DINAA, a major dataset and program participating in this project. The DAACS is an open source (Postgresql) database managed by a long-term consortium of institutional partners. In keeping with FAIR principles, tDAR, Open Context, and DAACS all use Creative Commons licenses and metadata to explicitly document reuse permissions and expectations. tDAR and Open Context both issue Digital Object Identifiers (DOIs) to facilitate data citation and provenance tracking (potentially significant for CARE purposes).

Datasets from tDAR, Open Context, and DAACS will be used to demonstrate FAIR+CARE principles and good practices. The primary "data management" need for this project will center on exploring, developing, testing, and promoting CARE related metadata. The project will implement practices advocated by LocalContexts.org (the authoritative curator of CARE metadata). In keeping with LocalContexts' recommendations, the project will leverage widely supported Dublin Core Terms metadata attributes (especially provenance, rights, rightsholder, license, and source) and reference LocalContexts curated URIs to CARE related metadata concepts (TK labels, Notices, and Licenses). These metadata will be visible in the tDAR, Open Context, and DAACS user interfaces and will be expressed using existing APIs and Webservices (JSON-LD, OAI-PMH, etc.) offered by these platforms.

Un/Semi-Structured Data

The project will develop policy guidelines and instructional materials that will take the form of static digital documents (PDFs), presentations, videos, and possibly podcasts (audio files). All these materials will be made available in nonproprietary open formats. They will be archived in tDAR and Zenodo, both of which will manage versioning, and will be documented using Dublin Core Terms metadata standards and will receive DOIs for citation purposes. We expect to publish most materials with the Creative Commons Attribution license, but some materials may be published under more restrictive licenses (including CARE related licenses) if required by consensus with stakeholders.

Dynamic (Software-driven) Content

The FAIR+CARE Cultural Heritage Data Assessment Tool will be a simple HTML and Javascript (with JSON configuration files) web-based tool to provide interactive guidance on FAIR+CARE implementations. To reduce costs and complexity, the implementation will be strictly "client side" without a server-side database or code dependencies. Open-source frameworks, like Bootstrap and Vue.js, will power dynamic user interface elements. The tool will be developed with Git public version control, using Github project management and collaboration tools. It will be deployed initially using Github pages with redirects from already purchased domains (FAIRarchaeology.org, FAIR-archaeology.org, and FAIR-CAREarchaeology.org will direct to this site). The assessment tool and code will be released open source (GPLv3 license).

Survey Data

Qualitative data will be collected via a web-based Qualtrics survey. The survey will be anonymous. No emails or other identifying information will be collected. Data analysis files will be named by the type of analysis and the date. We will use several processes to ensure data quality, security, and confidentiality: 1) careful supervision of all staff handling data, 2) contemporaneous descriptions of data collection and analysis processes.

In compliance with IRB guidance, only aggregate data (of low risk for personal identification) will be made publicly available after analysis has concluded and the data has been prepared for (low risk) sharing. Peer-reviewed published articles which utilize the data and the final IMLS report will accompany the data. This will provide context for future users. Zenodo.org and / or the California Digital Library Merritt repository will provide archiving services for these data expressed in CSV tables. The Creative Commons Attribution license will be applied on these data, unless more restrictive licensing conditions are required by project stakeholders.

Unaggregated "raw data" of survey results will only be accessioned into an institutional repository if allowed by the IRB and if a repository with suitable security controls can be identified. If these conditions cannot be met, the raw data will be held for a minimum of 4 years after the conclusion of the project by project leads in offline storage as encrypted CSV files. Encryption keys will be shared among the project leads.

The Alexandria Archive Institute (Open Context)

Organizational Profile

Statement of Purpose: The Alexandria Archive Institute (AAI) was incorporated as a non-profit organization in the State of California on August 1, 2001 (Articles of Incorporation) with a purpose to enhance scholarship through use of the Open Web by providing the global scholarly and educational communities an online archive of world archaeological, historical, and environmental data. This goal led to the development of Open Context, a web-based publishing service for archaeology and cultural heritage data. In 2018, the AAI began doing business as Open Context and legally operates under both names.

Governance: The AAI is governed by a nine-member board of directors; Open Context has an editorial board comprising 15 domain experts. The AAI has a staff of five individuals who collaborate on all AAI projects in various capacities. The Executive Director manages the organization's daily operations. (https://alexandriaarchive.org/about/people/). Our projects build expertise and capacity for innovation across the humanities. To ensure longevity and long-term citability, we archive data with the University of California's California Digital Library (CDL) Merritt repository and Zenodo, and use the Internet Archive for image hosting.

Communities Served / Audience: The AAI's work has a global reach, serving communities interested in archaeology and cultural heritage. We follow the American Library Association's privacy guidance by not tracking information about individual Open Context users, so we do not have a precise understanding of user demographics broadly. The proposed project focuses on networking with diverse communities of practitioners, including cultural resource management and Tribal professionals, and museums, libraries, instructors, and researchers. Over 20 years, the AAI has built relationships with researchers and other professionals across these sectors, including collaborations on multi-year research projects. Recent projects include the Secret Life of Data (SLO-data) Project[1], a longitudinal study of practices of creation, management, and re-use of archaeological data drawn from three geographical areas to investigate data quality and modeling requirements for re-use by a larger research community; and a Data Literacy Program developing resources to widen and diversify engagement with cultural heritage data among educators, professionals, and the public.

History: Since 2003, the AAI has secured continual funding from government grants, private foundations, individuals, and consulting. The AAI's work focuses on the open access data publishing system, Open Context (https://opencontext.org). Open Context is an established and globally recognized service for the editorial review, curation, and publication of archaeological and related data. The National Science Foundation (NSF), the National Endowment for the Humanities (NEH), and the US Department of State (Cultural Heritage Center)[2] reference Open Context for grant data management. Open Context and its developers have received awards from the International Digital Curation Conference (2014), the Archaeological Institute of America (2016), and the White House (a "Champion of Change" award in 2013). Open Context's data publications include 180 projects worldwide, representing 1200+ researchers and more than 2 million records, including 140,000 media items. Significant datasets include UNESCO World Heritage sites (Petra, Catalhöyük, Giza) and the Digital Index of North American Archaeology (DINAA), the most comprehensive set of data documenting ancient settlement in North America, currently with 850,000 site records. The AAI also maintains smaller collections, such as a collection of modern art lost and looted in the aftermath of the Iraq War and a geospatial visualization of Classical texts. (https://alexandriaarchive.org/about/history/).

[1] See: https://www.oclc.org/research/areas/research-collections/slo-data.html

[2] See: https://eca.state.gov/cultural-heritage-center/resources