NATIONAL ENDOWMENT FOR THE
Humanities

OFFICE OF **CHALLENGE GRANTS**

Narrative Section of a Successful Application

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Infrastructure and Capacity Building application guidelines at www.neh.gov/grants/preservation/infrastructure-and-capacity-building-challenge-grants for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Challenge Grants staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title: Expanding and Sustaining an Open Future for the Past: Data Literacy and Community-Building in Digital Heritage

Institution: Alexandria Archive Institute, Inc.

Project Director: Sarah Whitcher Kansa

Grant Program: Infrastructure and Capacity Building Challenge Grants

3. NARRATIVE

The Alexandria Archive Institute (AAI) seeks a National Endowment for the Humanities Infrastructure and Capacity Building Challenge Grant for a project with the following elements: (1) expansion of our core archaeological data publishing and archiving services; (2) development of a data literacy program for broader public education; and (3) establishment of a consortium of museum, library, and other institutional sponsors to sustain open access to archaeological data into the future.

The challenges inherent in understanding and using digital data are as intellectually demanding as any other humanistic research endeavor. Confronting these challenges requires dependable institutional and financial support. The AAI has a 15-year record of success publishing major data data sets, books, and peer-reviewed articles. The AAI's staff holds leadership positions in scholarly societies, policy-making committees, and granting programs, advocating change in professional attitudes towards data. In 2013, the White House celebrated our contributions with a "Champion of Change Award", and we subsequently earned awards from the digital library and archaeological research communities. Even with these tremendous strides, our operations as an independent non-profit organization rely on tenuous, high-risk grant support. To build a stronger foundation for continuity and growth, our Challenge Grant project will strengthen and diversify broader community engagement and institutional partnerships.

3.1 SIGNIFICANCE AND INTELLECTUAL QUALITY

The AAI works to transform data into knowledge, making them valuable and relevant to scholars and the public. We do not archive research papers, but instead focus on primary-source digital content (databases, photographs, field-notes) that archaeologists use to build arguments and interpretations. Thus, we do not replicate the role of book or journal publishers or repositories of digitized literature (JSTOR, the HathiTrust). Rather, we offer new services that enrich and complement existing publishing with vast bodies of underlying primary data. By publishing and archiving results of often destructive research methods, particularly excavation, we help meet critical needs in cultural heritage preservation. Furthermore, since the AAI's inception, we have researched, employed, and promoted ethically and

contextually-appropriate approaches to open access and open data (Kansa 2009, Kansa 2016, Kansa et al. 2005; also see **section 10.8.3** and **9.2 Letters of Support**).

Our data publishing service Open Context (http://opencontext.org) emphasizes intellectual investment in data. Data preservation cannot be regarded as a simple compliance issue. Data involve significant ethical, including cross-cultural intellectual property issues, modeling, and conceptual challenges (see **section 10.8**). For data to be meaningfully preserved and used in intellectually rigorous ways, they need to be integrated fully into all aspects of professional practice, including ethics, teaching, and publishing. To succeed over the long term, we need to strengthen the human and community capacity to use data effectively and appropriately. Thus, our proposed Data Literacy Program and consortium will help expand and strengthen our community of users, partners, and supporters. The following subsections describe each of these three pillars of our proposal.

3.1.1 Expansion of Core Activities

Open Context publishes a wide variety of open archaeological data, from archaeological survey data to excavation documentation, artifact analyses, chemical analyses, and detailed descriptions of biological remains from archaeological contexts. We use 100% open source technologies to ensure control over critical infrastructure. By integrating distributed information services (especially university library archiving), we have built a scalable platform for archaeological data publication. The NSF archaeology program and the NEH Office of Digital Humanities reference Open Context for grant data management.

Our scale is on par with those of major museums. Open Context's data publications include 122 projects worldwide (including forthcoming), representing over 800 researchers who, combined, contributed over 125,000 images and 1.5 million records from all inhabited continents. Publications include UNESCO World Heritage sites (Petra, Catalhöyük, Great Sphinx of Giza) and the Digital Index of North American Archaeology (DINAA) project, the largest and most comprehensive set of data documenting ancient settlement in North America, currently with 500,000 site records. The range, scale, and diversity of these data require expertise in data modeling and a commitment to continual development and iterative problem solving (see **section 10.5** for more information on this diversity).

In general, digital repositories inside (Huggett 2015) and outside of archaeology (Faniel and Yakel 2017) face challenges in encouraging use of their collections. Open Context addresses reuse challenges by working toward integration, interoperability and contextualization of information. Our publication that describes Open Context's editorial practices and implementation of Linked Open Data for data integration (Kansa et al. 2014), won wide interdisciplinary recognition with an award at the 2014 International Data Curation conference. Our approach integrates multiple disparate archaeological datasets, including government-created archaeological site inventories in the Digital Index of North American Archaeology (DINAA) project (Anderson et al. 2017). Some of these data have already supported multiple publications by different teams of researchers (Arbuckle et al. 2014; Atici et al. 2017) as well as instructional use (Cook et al. 2018). As outlined in **section 10.7.1**, various Web and citation metrics also demonstrate Open Context's wide reach and impact. Our Letters of Support (**section 9.2**) also highlight significant impacts.

Researchers will invest time in learning how to use a new information resource if it has sufficient relevance and significance to them. Thus, to broaden and deepen our scholarly impact, Open Context will continue to publish and integrate high-quality data, as well as study research data management workflows and user needs. Our current NEH-funded study exploring data creation and reuse patterns informs our strategic planning and guides improvements to our publication services and user interfaces.

3.1.2 New Program in Data Literacy

Data and algorithms shape the actions of virtually every institution in modern society. They help determine where police focus enforcement efforts, who gets health care, who gets welfare benefits, and who gets targeted for propaganda (O'Neil 2016; Eubanks 2018). Yet, such software-encoded decision making rarely sees wider scrutiny or accountability. Uncritical hype (see Morozov 2013) surrounding the term "big data" highlights the need to bring humanistic traditions of critique and wider cultural and historical perspectives to public conversations about data. To encourage a more just civil society, we need to harness these humanistic understandings to recognize and document implicit biases in using data. Doing so will help insure that data serves broader and more inclusive needs.

As highlighted by Killgrove (2018) in a recent web seminar, movements such as the March for Science, which first occurred in 2017 at over 600 cities worldwide, highlight broad public interest in evidence-based policy-making. The reaction to our 2017 *PLOS ONE* paper (Anderson et al. 2017) confirms Killgrove's points. That paper described climate-change driven coastal flooding threats to tens of thousands of archaeological sites documented by DINAA (data hosted by Open Context). The paper earned 25,000 views and widespread press, with coverage in the *Washington Post*, *Newsweek*, *The Guardian*, *National Geographic*, *Forbes*, *Wired,* and more. These articles all highlighted how access to public data informed the analyses.

A survey conducted by the Society for American Archaeology (SAA 2000) aimed at understanding how the American public learns about archaeology, and found they engage with archaeology through journalism and documentaries delivered via mass media such as magazines and television (which today would include social media and the Web). To reach a wide audience, therefore, our proposed Data Literacy Program will combine elements of *National Geographic* style journalism with the kinds of "data journalism" pioneered by Nate Silver and the *FiveThirtyEight.com* website.

The Data Literacy Program will develop narratives (including articles, videos, and data visualizations) intended for undergraduate and lifelong learners, that draw on archaeological (and other) data. Archaeology itself is a uniquely-data rich humanistic discipline with untapped potential to promote broader data literacy. Archaeological data, when explored with other publicly-accessible historical, environmental, and sociological datasets, offers endless opportunities for exploration. Articles will open doors to multiple levels of engagement, ranging from more casual readings of research questions and results, to more in-depth exploration of data analysis methods and primary datasets. Making documented code and data accessible will promote more transparent and accountable "reproducible research" practices (Marwick et al. 2017), thus contributing to culture changes with impacts inside and outside the Academy so that wider communities can peer into— and contest— processes that are now hidden inside analytic "black boxes."

The Data Literacy Program stories will be hosted open access on our own site or on venues that syndicate to major popular science magazines. Our team will include a program director and editor with archaeology and journalism background, a technical director, and contracted content developers / writers (see **Section 8**). We will begin with a one-year pilot phase, during which time we will produce three stories and solicit feedback from users on their usability and engagement, and then build up to 12 stories per year for the full program.

3.1.3 Consortium Development for Sustainability

The Web makes open access and open data feasible by dramatically reducing dissemination and copying costs (Suber 2012:4). Even so, producing high-quality information requires investment and human expertise. Conventional business models cannot sustain open access publication of archaeological data. Only approximately 10% of our income comes from fees to publish materials online, and these fees impose high burdens on junior scholars and scholars lacking institutional or grant support. To more equitably finance our operations, we need to cultivate partnerships with institutions that already purchase information services in support of research and teaching. For example, tapping into the acquisitions budgets of research libraries to support open data publication can help us be more sustainable.

Our sustainability strategy will focus on developing a consortium of existing institutions already supporting archaeological scholarship and curation services. To achieve this goal, we will form a Sustainability Advisory Board, composed of representatives of university libraries, museums, and archaeological research institutions (see **section 7.2** for a list of members and their letters of commitment in **section 9.1**). Their guidance will help shape how field research can integrate with museum and library information systems. They will also help us develop strategies to build financing partnerships.

A consortium of sponsors best reflects our commitment to providing public information goods that benefit the whole community. Our key challenge in developing a consortium of support centers on overcoming "free-rider" problems (benefiting without contributing). Fortunately, other open access publishers have already pioneered a consortium model. Our closest analog would be the *Open Library of the Humanities* (https://www.openlibhums.org/) that supports open access publication of humanities

literature (Kennison and Norberg 2014). To enable longer-term growth, we will also explore coupling consortium support with delivery of technology services. A valuable model comes from the for-profit, open access publisher, *Ubiquity Press* (https://www.ubiquitypress.com/). *Ubiquity* developed a software platform for production, editing, and review of academic papers, and sells use of that platform to other publishers (mainly university presses), thus helping those presses reduce their operating costs.

NEH Challenge Grant funding will be used to fully develop a consortium business model that can incrementally grow while providing more predictable income. Our advisory board of representatives from libraries, museums, and publishers will work in conjunction with strategy consultants to identify specific budget, marketing, growth-targets and partnership policies. The network of partners can provide us with invaluable insights into their needs and ways we can collaborate to cut costs and improve data management and collections management services. Over the longer term, we will leverage a network of consortium partnerships to explore opportunities in Platform-as-a-Service and related business models. Individual universities usually cannot justify hiring in-house expertise in a specialized area like archaeological data curation. Therefore, if our sustainability efforts are successful, we can pioneer a model for wider adoption in other domains similarly needing specialized data curation services.

3.2 Long-Range Plans

Archaeology's digital data needs will only grow in the coming years and decades. The profession will continue to need organizations like the AAI, dedicated to research, development, and advocacy in advancing new forms of scholarly communications. Thus, our long-term core activities will continue to develop and expand Open Context as a leading publisher of editorially-vetted, open archaeological data. In promoting intellectual investment in data, we also promote greater reflection on the ethics and social context of data in archaeology (see **section 10.8.3**) as well as how data articulate with method and theory. We understand the need for wide community engagement with the challenges and opportunities offered by research data. In this way, sustainability involves more than financial issues. We also must cultivate vibrant communities of practice (Nowviskie 2012). Thus, the Data Literacy Program will help to broaden

and diversify intellectual engagement with research data and advance this field in exciting and unanticipated new directions.

Grant funds will support the Sustainability Advisory Board, composed of individuals from museums and libraries, as well as publishers and for-profit businesses, who are committing their expertise to solidify a plan for sustainability beyond the duration of the proposed project (see **section 7.2** and **9.1 Letters of Commitment**). Leveraging partnerships with organizations already invested with intellectual and social capital will extend the impact of our work. We already have established collaborations with archaeology's leading institutions and digital libraries, and now work with several publishers to develop policies and procedures for linking underlying datasets to the publications they inform (see **section 10.7.2**). We also have a commitment for [INSERT HOURS] of pro bono consulting services (valued at [INSERT AMOUNT]) in Year 1 for sustainability strategy consulting (see **letter of commitment**).

3.3 IMPACT

The challenges associated with data are vast, and the landscape of the digital humanities is constantly shifting. For example, Digital Antiquity, an allied institution with a different, yet complementary, model for digital data management, takes a more conventional approach, emphasizing data preservation with the tDAR repository. In contrast, Open Context's publication process emphasizes review, editing, exhibition and linking of dynamic data, not the preservation of static files (see **section 10.8** and **Figure 10**; for a librarian's comparative perspective of tDAR and Open Context, see Sheehan 2015). tDAR's services meet an important set of preservation needs, especially in cultural heritage management. However, neither tDAR, nor Open Context, nor any other single system, such as the Digital Archaeological Archive of Comparative Slavery (DAACS), can meet all needs of a community as wide and diverse as archaeology. We will continue to promote interoperability and collaboration across diverse systems. For example, Open Context already cross-references with systems inside and outside of archaeology, including tDAR, Pelagios (humanities), and VertNet (biogeography). We are working toward Linked Open Data cross-references with DAACS, the Chaco Research Archive, and others (Anderson 2017).

Moving forward, our new program to cultivate broader data literacy will provide much-needed scaffolding to guide both professional and student communities in thoughtful engagement with research data. Finally, our work to develop a sustainability model will benefit programs beyond Open Context. Many disciplines need data curation services, and typically an individual university cannot justify hiring in-house expertise to meet these specialized needs. Our efforts to develop a cross-institutional consortium can pioneer a model that other specialized curation services can adopt.

3.3.1 Planned Challenge Grant Expenditures

We seek combined funds totaling $2 million, which we will allocate over ten years in the following ways: (1) $110,000 annually for two leadership research positions for our core programs; $125,000 annually for the new Data Literacy Program; $16,000 annually for the Sustainability Advisory Board and development of a consortium; and $5,540 annually for data archiving needs.

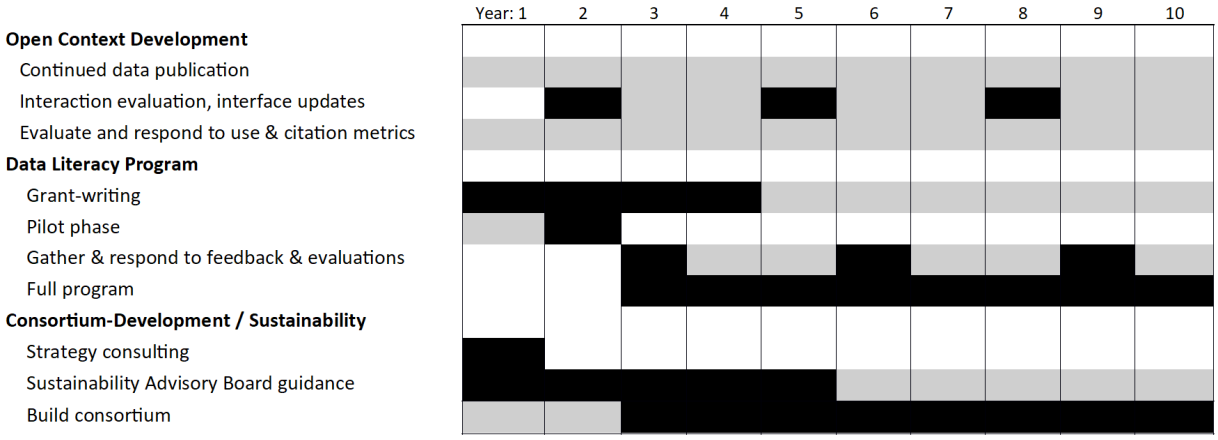
3.3.2 Plan of Work and Assessment

The assessment of scholarly endeavors represents a fundamental challenge, especially since major impacts may be diffuse and difficult to quantify. Nevertheless, certain benchmarks can help gauge the effectiveness of our programs and offer feedback for improvement. Our work plan (Table 1) includes phases for periodic assessment of our programs. To better protect user privacy, Open Context and other AAI systems do not impose login barriers. This complicates gathering some Web metrics because bots often mimic human users. Therefore, we will focus most assessment attention on evidence for reuse of content we publish, rather than simply quantifying views or hits. Assessment criteria will include:

- **Data Citation:** As data citation (facilitated by stable identifiers, especially DOIs) becomes more routine, we can track citations of Open Context data publications. Similarly, we will track citations and other metrics of articles and books that synthesize Open Context published data.
- **Software and Online Reuse:** Open Context data is also used in a variety of software applications (including virtual reality, augmented reality and mapping applications) and other information systems. We will identify and track such reuse through referring links.

- **Syllabus References:** We will gather data on use of Open Context and Data Literacy Program articles in course syllabi. Instructor feedback (via email) will help guide improvements.
- **Version Control:** Open Context software and Data Literacy Program code snippets will be hosted on GitHub, where issue tracking and commenting will gather feedback and guide improvements.
- **Consortium Partnerships:** Our sustainability goals can be measured by the number of institutional partnerships and the proportion of our operating budget that comes from partnership support.

Table 1: *Plan of Work, Showing Activities by Year (black = intense activity; gray = less intense activity)*



people in rural communities often access archaeological data, particularly DINAA, that describe cultural heritage resources in their vicinity. The Data Literacy Program will broaden and diversify this audience to include undergraduate students, their instructors needing guidance on teaching with data, and life-long learners interested in history, archaeology and related fields.

3.5 Plans for Raising Funds

The AAI's organizational independence permits a lean and efficient structure free of high indirect costs and large administrative overhead of most universities. Over the past fifteen years, the AAI has raised more than $2 million in funding from public and private sources, in addition to donations from individuals and fees for data publishing services (see **10.3 Projects and Funding**). Our fundraising plan centers on securing large grants from foundations and individuals that have programs, or have

demonstrated prior support for, archaeology and the use of technology in education. Our proposed Data Literacy Program is the key element to our fundraising plan. It will enable us to broaden our base of support to include funders of education, such as [INSERT NAMES OF POTENTIAL AND PAST SUPPORTERS]. We have strong ties with [INSERT NAME OF FOUNDATIONS], which already make annual contributions to the AAI to support our programs.

In anticipation of expanding our programs, over the past two years we have strengthened our board of directors to include individuals with contacts in the technology sector and private foundations that support educational technology and heritage. We have developed pilot plans for our Data Literacy Program and have applied for funding to launch the program in summer of 2018. We have received a commitment of *pro bono* support strategy consulting in Year 1 to help develop our sustainability strategy, valued at [INSERT AMOUNT]. The strategy team will work closely with our Sustainability Advisory Committee. We have also assembled a Fundraising Advisory Board of leaders with proven track records of fundraising success (see **section 7.3**).

3.6 Recent NEH Support

Since 2008, the AAI has received six grants from the NEH, totaling $900,000. These projects are listed at the NEH website (http://tinyurl.com/qhq3jlu) and more details are in **section 10.3**. We currently hold a Research and Development grant (PR-234235-16) to better align data creation and reuse. This project included a $25,000 match, which we secured in the first year from individual donations.