Fostering Collaborative Breakthroughs In Heritage Science via Machine Learning and Data Science

An Experts Meeting at the Intersection of AI and Heritage Science
Funded by the National Science Foundation
(IIS 2035533) [1]

Meeting 1: July 13-16, 2022, Hosted by The Getty Conservation Institute, Los Angeles
Meeting 2: June 19-20, 2023, Hosted by the Observatory of Heritage at Sorbonne University (OPUS) and the Sorbonne Center for Artificial Intelligence (SCAI)

In Partnership with

EduceLab and the University of Kentucky
The Getty Conservation Institute
The United Kingdom Arts and Humanities Research Council
The Observatory of Heritage (OPUS), Sorbonne University
The Sorbonne Center for Artificial Intelligence (SCAI)

EduceLab | A DIGITAL RESTORATION INITIATIVE

Arts and Humanities Research Council

Getty Conservation Institute

---

[1] Organized by W. Brent Seales (University of Kentucky) and Tom Learner (Getty Conservation Institute) and their respective research teams, these meetings were generously funded through a grant awarded by the National Science Foundation (IIS 2035533). In addition to the support of the NSF, we gratefully acknowledge sponsorship from the Getty Conservation Institute, the Arts and Humanities Research Council of Great Britain, EduceLab at the University of Kentucky, The Observatory of Heritage (OPUS) at the Sorbonne, the Sorbonne Center for Artificial Intelligence (SCAI), and Mechdyne (mechdyne.com).

Meeting Co-Organizers

W. Brent Seales
Professor of Computer Science
University of Kentucky

Tom Learner
Head of Science
Getty Conservation Institute

Organizing Committee

Chance Coughenour
Program Manager
Google Arts and Culture

Gregory Heyworth
Professor of English and Textual Science
University of Rochester

Sylvester Johnson
Professor and Director of the Center for Humanities
Virginia Tech

Holly Rushmeier
Professor of Computer Science
Yale University

Walter Scheirer
Professor of Computer Science and Engineering
University of Notre Dame

Steve Seitz
Professor of Computer Science and Engineering
University of Washington

Jeanne Marie Teutonico
Associate director of Strategic Initiatives and Publications
Getty Conservation Institute

Meeting 1 Participants (The Getty, Los Angeles): Taylor Berg-Kirkpatrick, Dominique Bouchard, James Brusuelas, Christy Chapman, Vincent Christlein, Doug Comer, Simon Crutchley, Fiona Dakin, John Delaney, Deena Engel, Seren Griffiths, Mark Griswold, Mami Hayashida, Lise Jaillant, Henry Kautz, Daniel Keefe, Ira Kemelmacher, Joe Kider, Doug MacLennan, Stefania Merlo, Paul Messier, David Messinger, Mark Mudge, David Newbury, Stephen Parsons, Patricia Murrieta-Flores, Michael Phelps, Lilly Pregill, Emma Richardson, Mario Santana Quintero, Carla Schroer, Liz Smith, Todd Swanson, Corey Toler-Franklin, Giovanni Verri, Claire Warwick.

Meeting 2 Participants (The Sorbonne, Paris): Emma-Jane Alexander, Frédérique Andry-Cazin, Gérard Biau, Christy Chapman, Vincent Christlein, Marzia D’Angelo, Louis Falissard, Nathalie Ginoux, Chloé Guennou, Lise Jaillant, Henry Kautz, Nicolas Leys, Stefania Merlo, Tsveta Miteva, Federica Nicolardi, Seth Parker, Stephen Parsons, Emeline Pouyet, Gary Quasebarth, Corey Toler-Franklin, Josef Wilczek.

Table of Contents

1. Executive Summary

This report summarizes the results from two meetings organized with partial support from a grant awarded by the National Science Foundation (IIS 2035533). The first meeting was organized by W. Brent Seales (University of Kentucky) and Tom Learner (Getty Conservation Institute) with the help of their respective research teams and the organizing committee. The second meeting was organized by W. Brent Seales and Frédérique Andry-Cazin (The Observatory of Heritage (OPUS), Sorbonne University). The two meetings covered a wide-ranging set of topics around the theme of AI and heritage science and included the diverse voices of participants from many different backgrounds. The meetings successfully elicited in-depth conversations pointing to consensus trends and issues from experts and practitioners with experience and vision[2]. This report synthesizes the interactions and conversations around those issues and organizes them into a discussion of opportunities, major challenges, solutions, and broader impacts.

As the following sections will detail, the workshop discussions were wide-ranging and reflective of a diverse set of the participants' interests and experiences. With a high degree of consensus, workshop participants asserted the following touchstones around the practicalities of progress and impact:

1. **True collaboration across disciplines** is key to advancement of the field, especially when considering the rise of AI and Data Science as applied to Heritage Science. There are barriers that can be identified and systematically addressed.
2. **Image- and text-based collections** offer enormous opportunities for immediate advances, aligned with advances in AI, that will connect and challenge diverse fields of study and will also engage the public.
3. The **challenge of acquiring, managing, stewarding, and storing data** is a universal, practical behemoth threatening and inhibiting progress.
4. **Developing funding streams and programs** for projects at the confluence of these areas should be a top priority for the NSF, NEH, and various philanthropic foundations.
5. The **potential for positive, wide-ranging, durable broader impacts** is enormous and this moment for realizing those impacts must not be squandered.

The participants in these workshops agreed that there is now strong potential for significant and durable broader impacts that can engage researchers, push multiple fields forward in their respective approaches and advancements, and capture the public's imagination. The barriers are surmountable, and the potential return on investment is high. The time is now for intentionality around deep and sustained financial and programmatic investment to develop projects, provide career advancement and training for individual researchers, and deploy broader initiatives at the confluence of AI and heritage science.

[2] Meeting agendas and participants are listed in the Appendix.

2. Introduction

> "We are at a turning-point. Science and technology present a host of exciting opportunities to the heritage sector. They must not be wasted."
> *— UK House of Lords Paper 256,* Science and Heritage*, Report of the Science and Technology Committee, 2005-06.*

The recent convergence of several trends in information technology makes it possible to frame new challenges in natural and cultural heritage – and reframe old ones – through the potent "lens" of artificial intelligence (AI), data science, and machine learning (ML). It seems reasonable to expect that, armed with renewed vigor and remarkable technical breakthroughs, researchers will make durable and substantial advances in the interdisciplinary field of heritage science. However, new discoveries are unlikely to occur without intention and cooperation on the part of experts and funders from all sides of the research community.

In the summers of 2022 and 2023, computer scientists along with heritage science professionals and scholars convened for "Fostering Collaborative Breakthroughs in Heritage Science via Machine Learning and Data Science," a pair of intensive workshops sponsored by the National Science Foundation. Representatives from GLAM institutions (galleries, libraries, archives, and museums), universities, industry, and funding agencies convened to explore how the application of AI and data science could enhance research and conservation efforts in cultural heritage. The discussions among participants at the meetings revealed a nascent but eager community of scholars brimming with ideas, but facing funding and interdisciplinary collaborative challenges.

The organizers wanted these meetings to be the next exciting step in the continuation of a long-standing trend: the innovation and impact of computation, data, and now AI/ML in approaching challenges in the humanities. Well before the invention and availability of a computational approach, researchers and practitioners imagined and employed science and technology in the service of the humanities. Vannevar Bush envisioned that all the 20th century postwar science and engineering advances should be turned toward problems in knowledge creation and management, creating essentially a new kind of library. He used the bow and arrow as an example in his famous "Memex" paper[3]: "The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow." Bush was suggesting that the science of the Memex could facilitate a *humanities* project: exploring the history of and properties around human-constructed artifacts.

Bush stood at the threshold of the coming digital age. In the 1950s, Italian Jesuit Father Roberta Busa employed IBM accounting machines and punch cards to create his audaciously comprehensive concordance of the works of St. Thomas Aquinas. Without the IBM technology, it would have taken an estimated 50 scholars as much as four

---

[3] Vannevar Bush: "As We May Think", The Atlantic, July 1945. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/

decades to index the 13 million-word corpus[4]. This was arguably the genesis of so-called “digital humanities.”

The ability for applied science and technology to invent tools that do more with less is an obvious hallmark. The application of those tools to problems in heritage science has a rich history, and today's new technologies (chemical and elemental analysis; AI/ML/data science, advanced imaging of all kinds; large-scale datasets) provide much more than mere efficiency gains. Heritage science and the humanities arguably stand to gain the most from these new and remarkable technical advances.

For example, non-invasive techniques have exploded in capability and power. From materials analysis using Raman and x-ray fluorescence, to imaging methods based on spectral bands from ultraviolet to infrared, collecting data has never been more non-invasive and safe. And that makes an enormous difference: heritage studies is a field full of friable objects worthy of study but zealously protected from well-meaning – albeit probing – scholarly hands. Non-invasive data collection followed by digitization provides anyone with a computer the opportunity to study these materials safely and from a distance with no risk of damage. Even more revolutionary are advanced technical approaches that can become a path for a “digital restoration” of what has been lost due to time, weather, and human catastrophe without alteration of the original physical object.

Beyond materials characterization and advanced imaging, this landscape of opportunity has become astonishingly fertile with the recent maturation of Artificial Intelligence (AI). Previously unthinkable or unsolvable challenges in cultural heritage are now approachable with these new tools. In addition, complementary benefits accrue to the technological development community, as the heritage arena provides a unique and fruitful environment within which to develop increasingly sophisticated AI techniques that are sure to achieve wider application than in the heritage sector alone. For example, the latest and greatest AI technologies were born out of the humanities domain, yet have achieved widespread usage in a variety of fields. ChatGPT[5] is a large language model (LLM) that only works because of its voracious ingestion of humanity’s digitized written words, while DALL-E[6] and other image generator models like it[7] mine massive corpora of digital images to create a variety of “artworks” from textual descriptions, thus combining language and art in one high-powered tool. A productive symbiotic relationship between high tech and the humanities has emerged as the words and images manifested by our common cultural heritage have driven research and discovery in AI and machine learning environments.

The workshops we conducted sought to encourage this symbiosis by interfacing technologists with humanists to promote durable advances in AI-enabled heritage science. In total more than fifty studies deploying or suggesting AI applications to

[4] https://www.ibm.com/history/dead-sea-scrolls
[5] https://openai.com/chatgpt/
[6] https://openai.com/index/dall-e-3/
[7] https://stability.ai/

heritage collections were presented at the two workshops. The group's goal was not only to define specific opportunities around collections of artifacts, but also to surface the specific questions the artifacts raise, the challenges they represent, and the solutions that could be offered by non-invasive imaging, data collection, and artificial intelligence approaches. It was anticipated that specific heritage questions or problems would be matched with an AI technique to solve it, and collaborations between computer scientists and heritage scholars would naturally emerge as experts in the problem and experts in the solution found each other.

Viewed in the long timeline of the confluence of technology and the humanities, the workshops were an important, even seminal next step as AI finds its way into researchers' toolboxes. More specifically, however, the discussions revealed that this burgeoning, protean community remains quite nascent and unorganized. As noted earlier, cultural heritage scholars and professionals have been organically applying the tools and techniques of science and technology – a field known as heritage science – for some time. But the more recent deployment of data science, artificial intelligence, imaging science, and visualization techniques remains largely unsystematic, or even "messy."

For example, technology researchers seek to form compelling, important lines of inquiry around cultural collections, but they often lack a thorough knowledge of the objects, culture, and practices of the heritage community, which stymies their creativity. On the other hand, cultural heritage researchers intuitively and enthusiastically employ the complex state-of-the-art tools available to them, but they possess limited technical knowledge, a situation that creates a gap between full potential and praxis. This context is further complicated by the lack of a common language and differences in philosophical approaches, factors that all too often result in groups working in parallel with one another rather than in combination, slowing progress and innovation. The emergent nature of the application of AI tools to the study of heritage is not because the work lacks importance or promise, but because the approaches are scattered and "reinventing of the wheel" too often occurs.

Many of these roadblocks to progress have been mentioned by other heritage science study groups[8], some almost two decades ago: "We are at a turning point," stated the 2005-06 British House of Lords Science and Technology Committee report. "Science and technology present a host of exciting opportunities to the heritage sector. They must not be wasted."[9] With the emergence of AI onto the scene, the time is now to finally develop the robust mechanisms needed to encourage and nourish collaborative efforts and initiatives between the two communities of experts.

Against the backdrop of a long-term science and humanities tradition and the organic,

---

[8] Loïc Bertrand, Demetrios Anglos, Marta Castillejo, Bénédicte Charbonnel, Sophie David, et al.. D.9.3 E-RIHS Scientific Strategy v. 1.0. [Research Report] European Commission. 2020. hal-02946664
[9] HOUSE OF LORDS Science and Technology Committee 9th Report of Session 2005–06 Science and Heritage Report with Evidence Ordered to be printed 19 October 2006 and published 16 November 2006 Published by the Authority of the House of Lords London :HL Paper 256

rapidly changing emergence of powerful AI tools and approaches, we present the opportunities, major challenges, solutions, and broader impacts that the workshop attendees discussed and identified as significant and compelling for future research potential and advancement. The time indeed remains “now” to plan, invest, and reap the advances to come from the fostering of collaborations in this arena.

3. Opportunities

The galleries, libraries, archives, and museums (GLAM) universe is uniquely equipped to both enable AI development and benefit from it. All AI and machine learning systems are powered by data – massive quantities of data. And the vast network of GLAM institutions have access to a plethora of high quality data that can feed the development of new AI applications and, in turn, power incredible discovery about our cultural and natural heritage. As one participant stated, “I have been surprised about the questions I have been able to answer due to the incredible scale of the data that is available and usable to these systems, questions that not long ago, I could not have even asked.” Another noted, “We are getting deep insight into humanity with this technology. We are able to understand things about ourselves that we couldn’t without it.”

There are two primary types of that data that all GLAMs possess: text and image. As the repositories of massive amounts of existing and potential textual and image data, GLAM institutions provide a fertile ground for significant breakthroughs in discovery using AI.

Textual Data

GLAM institutions maintain the world’s collections of texts, the vast amount of which can be used as training material to create large language models specific to the needs of cultural heritage scholars. This poses exciting opportunities:

- Extinct languages (e.g., Linear A, Minoan culture language) – Are there AI-inspired methods to find patterns in data that humans could never find?
- Advanced search and summarization – Can LLMs be trained to usefully report and summarize the content and the nuance of the many disparate textual collections in libraries all over the world?
- Lacunae and new textual discoveries – Can new methods lead to the discovery of new texts, new archaeological sites, and new information embedded in objects that are otherwise impenetrable?

GLAMS also generate textual information of their own that can be harnessed for discovery. For example, by synthesizing all of the information from conservation reports across heritage institutions into one textual corpus, a natural-language-searchable “common body of knowledge” can be created. Such tools can be used to facilitate decision making by giving conservators access to the minds and activities of all conservators across the globe. In addition, automated data analysis can help reveal patterns in how various collection types respond to certain preservation interventions. It can also uncover relationships between indoor and outdoor environments that enable predictions about internal change based on what is happening externally.

Image Data
As generative AI tools such as DALL-E and Midjourney have revealed, the expansiveness of image collections provides a deep well of training material for AI tools. And the number of images in humanity's catalog is immense. According to Photutorial, there are approximately 14.3 trillion photos in existence, and many of those remain in analog form, awaiting digitization. In addition:

- Globally, humans capture 5.3 billion photos daily, or 61,400 per second.
- 750 billion digital images are on the internet, and Google Image Search indexes an estimated 136 billion of them. 14 billion images are shared daily on social media.
- In 2024, an estimated 1.94 trillion photos will be taken worldwide.
- In 2025, humanity will take over 2 trillion photos in one year. That number will reach 3 trillion by 2030.

These stats do not include such material as digitized texts or 3D images.

This ocean of visual images is vast, but museums, libraries, and archives have access to those considered the most significant, and their collections increase daily. These image-based collections along with the geometric data that accompanies them offers valuable data for the development of AI applications in every heritage area. Some potential opportunities:

- The detection of fake images, fake news – Image data contains information and details that the human eye cannot possibly detect and integrate over large numbers of images. AI methods might be able to address the problem of inauthentic claims based on a summarization of the nuances in a large corpora of vetted photos with appropriate associated metadata.
- New understanding of image/artistic domains – Image data may be able to help infer material and aesthetic characteristics of artwork, giving experts more tools in this area.
- Numismatic applications – There are certain subtle aesthetics in the structure and appearance of coins that ML has the potential to reveal and classify that are difficult for humans to perceive.

Supplementary Data
The artifacts in GLAM collections can themselves serve as training data for AI systems. Heritage objects possess a material history that starts from the moment of their creation and continues as they are stored in environments all over the world. Each object becomes an instantiation of all of the different environments that have existed over time. In other words, the objects themselves serve as sensors that "collect" data in the form of evidence that can be measured and combined with historical data to glean insights.

Large scale resources are expended in order to responsibly curate and conserve important collections worldwide. Many decisions about how to allocate resources are made with less than perfect data; often decisions are made with almost no real certainty about the tradeoffs in environmental conditions across a wide variety of collections and preservation options. Some opportunities:

- Data-driven configuration of conservation environments – Historically the standards for maintaining conservation environments have been rather strict (21c, 55% RH eg). New research is suggesting that there could be more latitude, without increasing risks to collections, using an AI-based approach.
- The interplay between indoor and outdoor environments – AI might have the power to detect subtle relationships between indoor and outdoor environments, leading to better predictive power and data-driven collection management decisions.
- Comprehensive degradation analysis – It has been a long-term challenge to understand degradation/aging across the many parameters affecting collections (material, environmental, historical). AI-based frameworks may transform our ability to better understand and visualize this complex landscape.
- Decision-making for exterior sites – Patterns in LiDAR and other remote sensing tools open the door for AI/ML frameworks that could assist in predicting impacts for sites subjected to climate change, potential future excavations, and consideration for national registries and protections.

Skills Enhancement and Transfer

Collaboration is most robust when it leads to synergistic interplay and actual skills enhancement and transfer among participants, particularly interdisciplinary participants. David Epstein argues that leading researchers systematically seek to extend their "range" by understanding new concepts and areas of inquiry.[10] The opportunity exists at the confluence of AI and heritage science to foster and facilitate such skills enhancement and transfer of understanding.

Workshop attendees broke down the opportunities for skills transfer and transdisciplinary interchange into four primary areas: data, development, collaboration, and breakthrough. The data arena provides a natural point of overlap, since data acquisition usually requires the cooperation of multiple parties and creates an opportunity for cross-disciplinary skill enhancement during the process. Development is the intentional nurturing and fostering of skill through training. In the case of interdisciplinary development, such intentionality occurs through new programs, workshops, and tutorials designed for skills transfer; tools that democratize new methods, making them easier to learn and apply; and sustainability plans that ensure the availability of cross-training opportunities long-term. Collaboration involves developing actual policy frameworks and professional opportunities that incentivize the authentic occurrence of interdisciplinary partnerships and, more importantly, provides the structures that make them possible. Creating networks among people working on similar problems and providing a digital platform for sharing resources are two such structures mentioned by participants. Finally, breakthrough refers to the intentional

[10] David Epstein, *Range: Why Generalists Triumph in a Specialized World.* Macmillan, 2019.

expectation and preparation for outcomes with broader impacts that can be amplified through prizes, job creation, and public engagement.

4. Major Challenges

As counterpoint to the enthusiastic potential of impact and advancement, we asked workshop participants to inventory the most significant barriers and challenges they see to the realization of true progress. While any group of researchers and practitioners will recount the struggle with the quotidian, we worked hard to arrive at consensus on the unique barriers with AI and heritage science. The participants arrived at three primary barriers/obstacles: data and collection access and associated management issues; disciplinary divides; and resource limitations.

Too Much Data, Too Little Data

Although the world is literally drowning in data[11], it is not always simple to access the right kind of data. Even in the non-regulated world of what is available as open source and on the Internet, datasets are ill-structured, difficult to assemble, and non-standard in their usage rights agreements. The GLAM universe faces struggles to manage, access, and store data, – issues representing well-known problems common to every industry and organization.

The difference in the GLAM environment lies in the uniqueness of the data (and potential data) these institutions possess, the digital rights around that data, and its structure. While GLAM datasets are often carefully curated, which should represent a highly valuable source for AI development, the goal of the curation has not been to support AI analysis. Thus the data held and generated by GLAMs is well-structured for the purpose of the operations of individual projects and institutional goals, but not necessarily for the direct application of AI approaches. This barrier can be unusually resistant and requires resources to address.

Participants listed a number of related barriers revolving around GLAM data:

- Interoperability – It is often the case that the raw data formats and the linked open metadata formats that could power robust searches and algorithms are highly fragmented and do not interoperate well.
- Laser-focused, unrelated studies – One participant cited a common phenomenon where, for example, a focused study for a single object ends after very limited results and cannot be more broadly extended. Since those results do not get linked to other related studies, the archive fills with one-off results that miss the opportunity for broader collection-wide analysis. Since there is no focused, big-picture intention when creating a dataset, the institution as a whole loses this chance at broader impact.
- Multi-modal power – AI's strength lies in its ability to make multimodal connections and tease out nuances over disparate data. However, for studies seeking to connect multiple modalities, the aforementioned barriers to data

[11] By 2025, it is estimated that 463 exabytes will be created *daily* on the internet *alone* (Vuleta, 2021, p. 8)

access and acquisition are amplified. Although AI-enabled data fusion techniques should be more regularly explored, these barriers prevent it.

- Visualization – Because of multiple non-interoperable systems (and data that has not been organized into coherent packages), it is rarely possible to visualize the many different kinds of data collected on objects. One participant wondered if "there is a lot of gold in there we aren't even seeing." The barrier to more robust visualization, however, is structural (organization of data) as well as tool-based (which tools will perform the custom visualization and analysis that would be most helpful?)
- Differing notions of "scale"– In a world full of data, it is ironically the case that individual GLAM institutions often lack *enough* data to develop AI applications. The niche, highly focused nature of studies in the humanities often means that scholars are dealing with a very small set of materials. Yet AI methods repeatedly show that "more is better." Validation sets needed to train machine learning algorithms in these areas are viewed as luxuries due to such scarcity of data. But, participants noted, data is likely hidden across GLAM institutions that can be mined in innovative ways by AI to supplement existing datasets with new information that is not readily available or visible.

Disciplinary Divides

Workshop participants created a microcosm of the very disciplinary divides that form barriers for advancement and breakthroughs. One serious barrier when communities of practice that are different begin to intersect is the need for education. Humanists need to learn more about emerging technical approaches, and technical scholars need to learn about the questions, approaches, and constraints in the humanities and with GLAM collections. These barriers are real and need intentional strategies for overcoming them. We did not use workshop time to educate each other with tutorials and details about, for example, AI methods and conservation practices. But we clearly can see these differing practices and vocabularies as barriers to overcome.

The cultures of high-tech computer science and scholarly cultural heritage really are very different and have traditionally pursued divergent goals. Computer science as a field is marked by innovation at maximum speed. Using Moore's Law as an example, the exponential growth in transistors in an integrated circuit (doubles every two years) sets the tone for the kind of pace expected by the tech community. Admittedly these are generalizations, but work in many areas of humanities with GLAM collections is the opposite of speed and innovation. The expectation instead is systematic rigor and conservation at all costs of archives that are fragile and sensitive.

In their Open Digital Archaeology Textbook[12], the authors describe the schism in terms of Nassim Taleb's concepts of fragility and antifragility[13]. They note that humanists see something fragile as being likely to break under stress and randomness, a situation to obviously be avoided at all costs. Technologists on the other hand think of things as being "antifragile," whereby objects actually get stronger when exposed to stress and

---

[12] Graham et al., the Open Digital Archaeology Textbook. https://o-date.github.io/

[13] Nassim Taleb. *Antifragile: Things That Gain From Disorder.* Random House, 2012.

randomness:

> Academia’s systems are ‘fragile’ in that they do not tolerate failure; they are to a degree resilient, but they are not ‘antifragile’ in Taleb’s sense. … So silicon valley really means ‘fail’ in the sense of ‘antifragile’ but they frequently forget that academia sees ‘fail’ as the breaking of something fragile; and so the two are at loggerheads. (Graham et al., 2020, p.1)

In addition, heritage scholars generally focus on the broader human and cultural contexts of an issue, while the AI community zeroes in with laser focus on the computational problem and solution at hand. During the Fostering workshops, for example, ethical questions regarding the use of AI and cultural data appropriation were front and center of the heritage community’s mindset, while they were at best an afterthought in the collective thinking of the computer scientists. On the other hand, the technologists in the room grew frustrated as the articulation of specific technical needs faltered and the development of potential AI-enabled answers never materialized. As one participant noted:

> “I’ve had projects before where there were issues with translation between stakeholders, where we never achieved clarity regarding mutual objectives and other details. If you don’t know these things and agree on them, then you can’t make a clear plan of how to proceed.”

During the meetings, heritage professionals also expressed feeling threatened in the face of some of these hi-tech tools. As noted by one participant, “For a historian, if history were a person, this would be like sticking a knife into its heart and ripping out its soul and throwing it on the ground, stomping on it, and lighting it on fire.”

One example of this situation can be seen in the way the traditional role of the curator as mediator between a heritage object and a patron can be eroded by the entrance of technology into the relationship:

> “Technology in effect becomes the mediator, rather than the human curator whose role has been to rely on his or her expertise to facilitate how others experience heritage objects. With AI, technology can become as good or better at this, as machine learning can learn how humans implicitly understand or enjoy an object and use that knowledge to improve the patron experience.” (Workshop Participant)

Such very real disciplinary barriers create a risk of widening the gap rather than bridging it. It is all too easy to talk about interdisciplinary collaboration, while effective practice of that ideal often evades the best intentions. But the participants described experiences that proved incredibly fruitful when they engaged in systematic and truly collaborative efforts that bridged the divide by "learning our way forward" into new areas, new challenges, and new ways of working together.

One clear benefit from collaborative work is the translational opportunity of the diffusion of technical approaches into new fields. But diversity itself also offers advantages:

> "Fields like conservation and cultural heritage experience different demographic challenges than those common to computer science and engineering, and the collaboration between these divergent communities can together strengthen and inspire a broader, more diverse participation."
> (Workshop Participant)

In addition, funding challenges can be better addressed when collaborative researchers share the burden of developing research support across a number of fields rather than within the limited resources of only one or two disciplines.

It is therefore impossible to overstate the importance of "bridging scholarly cultures" and breaking down barriers in any discussion regarding the application of AI to the study of natural and cultural heritage. The field needs vocabularies and funding structures to bridge the gap between heritage studies and computer science and to allow new approaches to emerge and be adopted.

Resources

Ask any researcher what constitutes the number one impediment to their progress, and they will cite a lack of resources. However, the unique nature of heritage science as the fusion of high tech, hard science, and human culture creates a complicated environment to support. The study of heritage is highly interdisciplinary, and the lack of resources exacerbates the already gritty wheel of collaboration. Heritage projects also often have long life-cycles that rarely line up with funders' timelines of support. In addition, projects often require international teams. Such collaborations are expensive both in terms of cost and oversight/management, as funding opportunities are often siloed by country and limit opportunities for non-citizen scholars to participate in grants with their international colleagues.

Most importantly, the expertise and resources needed to develop and use AI tools in the heritage community is out of reach for most, a key point emphasized by workshop participants:

> "At some institutions like mine we can't even buy a computer. We can't even consider decisions between cloud computing or on premise servers or whatever because there is no knowledge or funding to support those necessities. The only way forward for us would be as a research partner, as there is just no way we can support the actual AI work. We small organizations have a lot to offer for AI/ML because we are more agile and can more easily collaborate, whereas larger organizations need so much time and effort to arrange participation with others."
> (Workshop Participant)

In addition, finding qualified human resources in this niche area is difficult[14]:

> "The emerging democratization of digital skills, methods and techniques in the digital humanities and the computational sciences is not happening at the same rate in archaeology and heritage. This has created a situation in which many archaeologists and heritage professionals are facing digital challenges they have little or no training to address. This has created an opportunity for initiatives that seek to provide students, scholars, and professionals with critical digital skills."

Compared to medicine or other industries, few people seem interested in working to integrate AI and cultural heritage. AI professionals can make more money elsewhere, while experts in heritage often view high-tech approaches to be incompatible with their scholarly culture and personal abilities. Thus, there is a true need for infrastructure, especially for enabling collaboration and training personnel:

> "How do we help small or even large institutions access the actual compute infrastructure they lack, thereby preventing access to any of these technical tools right now? They might have a dataset and know about some AI method but their computers run Windows 7, they have no compute power, and their personnel are not trained to use much less develop these tools. Both the lack of infrastructure and its adoption and maintenance represents a big speed bump to the community."
> (Workshop Participant)

Other participants cited the need for new generations of instruments that are purpose-built to scale up for wide application:

> "We need options that lower the entry point in terms of skill, experience, and cost so we can get to scale and make collecting these datasets a reality. It's time for simpler, less expensive, interoperable options that make up a more robust heritage data pipeline for ingesting, aggregating, and making available data across the entire community."
> (Workshop Participant)

U.S. funding of heritage science significantly lags that of other countries. For example, the European Union runs large synergistic programs annually (e.g., ERC Synergy) and other programs that target funding for heritage science. The National Science Foundation only rarely funds projects that fall into the category of heritage science. None of the NSF directorates own the breadth of what heritage science has become, and none address the current opportunity of AI/ML inspiring new breakthroughs. The National Endowment for the Humanities (NEH) has increased its footprint and is poised to continue, although it remains in constant turmoil over annual, politicized budget fights. That struggle continues despite clear studies demonstrating that heritage science offers broad benefits to society in general and to science in particular.

---

[14] E. Watrall. Building Scholars and Communities of Practice in Digital Heritage and Archaeology. Published online by Cambridge University Press, 22 May 2019.

Participants agreed that funding from national agencies and from private foundations would substantially change the likelihood for durable and sustainable collaborations to emerge, and for broader impacts to reach their enormous potential.

5. Solutions

The workshop participants summarized solution paths to overcome the major challenges discussed above according to five major themes: bespoke funding; improved tools; open data sets; small-scale rapid development projects; and training options. These themes individually and collectively address the major challenges, allowing for advancements and the realization of broader impacts (next section).

Bespoke Funding

Resources are a necessary (though not sufficient) condition for any kind of progress. There is a dearth of grant mechanisms that are specifically designed to facilitate collaborative heritage science projects, to develop human resources in the field, and to sustain the complex projects throughout their life cycles. Bespoke funding mechanisms will provide the resources needed to train interdisciplinary teams in methods designed to grease the gears; to provide sustainability so that projects can take the amount of time necessary to ensure translational results and positive outcomes; to provide sustainability through attracting and training scholars and professionals with a dedicated interest and in and knowledge of AI enabled heritage science. Some key observations around funding from the workshop participants:

- Funding should be sustained and designated for cross-cutting (interdisciplinary) advancement of AI and HS programmatically by the NSF and the NEH.
- The funding agents should develop specific categories of support for the application of AI technologies to cultural heritage problems, including support for workshops, conferences and virtual networking initiatives; small grants for exploratory research, the creation of data sets, post-doc positions, and exchange of researchers between institutions; larger grants for more complex, multi-year, collaborative research projects; support for research infrastructure and the creation of research centers.

Funding is crucial in order to feed the other solution paths below (tooling, data, short-term projects, and training opportunities). As noted previously, GLAMs possess an almost unimaginable amount of raw data in the form of digitized and undigitized heritage objects. In addition, research talent across the spectrum is available with capacity, especially in early-career researchers, to join a funded effort. Although there is a constant challenge in any field to make the case for sustained funding, the payoff here is unique.

Improved Tools

Participants noted the progress of various fields as indicated by the maturity of the tools available for research: computer vision (OpenCV[15]); software engineering (GitHub[16]); AI and machine learning (pytorch[17]; tensorflow[18]); medical imaging (ImageJ[19]; MIPAV[20]; Insight Toolkit[21]). Advances in tooling — specifically software tools — could provide substantial leadership and stability for programs designed to innovate new AI tools for heritage science. In the same vein, participants identified other tools and tool-based approaches that could be unifying and problem-solving forces for advancement:

- Software engineering around standardizing the installation, use, and interoperability of common tools and interfaces.
- Infrastructure that can operate in low-resource environments when required, such as field tools that don't require constant network connectivity to a server system. Many in-the-field applications require tools to be used offline and synchronized with a cloud-based system later.
- Open source software as a default, with the goal of ongoing support through philanthropic endowments, foundation support, and government investment.
- Dashboarding and orchestration tools for AI-based experiments, especially being conducted on GLAM data, to make the visualization and improvement of results easily shareable and repeatable.

Open Datasets that Challenge the State-of-the-Art

The creation of merged, multi-modal datasets released to the entire field has been a strategy used in other fields to stimulate advances (astrophysics; bio-engineering, artificial intelligence, robotics, etc.). The solution to the dearth of such datasets in AI/HS is to incentivize researchers as well as GLAMs through funding, appropriate licensing, and the promise of wide distribution/citation. Workshop participants discussed a number of organic processes to incentivize the creation of robust and challenging datasets:

- Organize a series of online meetings specifically designed to address the issue and produce the datasets and the dissemination plan to manage it. This would connect the people who are already working on data sharing with the specific challenges of heritage science and the broader community.
- Create a shared resource of a digital space/playground where researchers and practitioners can release shared datasets and let people who are working on algorithms play with the data.
- Engage with platforms such as Kaggle for contest-based work, allowing for data access but also broader didactic opportunities around the data, its origin, and its broader significance.

---

[15] https://opencv.org/
[16] https://github.com/
[17] https://pytorch.org/
[18] https://www.tensorflow.org/
[19] https://imagej.net/
[20] https://mipav.cit.nih.gov/
[21] https://itk.org/

- Consider branding opportunities from large GLAM institutions. Branded datasets (e.g., Life photograph archive, acquired by the Google Cultural Institute, now known as Google Arts and Culture[22])

Participants familiar with European initiatives noted the influence of Europeana[23] (linking data and museums), which is maturing and is a model working quite well as a place for resources dedicated to creating, accessing, and using content in the cultural and creative sectors. It is marked by support from the European Commission to create and disseminate content through two work strands: deployment of common European data space for cultural heritage; and grants to researchers and institutions specifically targeted for the work to create and disseminate high quality data (particularly 3D).

Small-Scale, Rapid Development Projects

The practical, organic experience of many of the workshop participants included small-scale “scouting” projects, or projects “on spec” as a way to overcome disciplinary barriers and promote social development[24]. With low-stakes projects that are small-scale and shorter term, it becomes possible to envision broader solutions and to build trans-disciplinary tools and approaches. Many participants pointed to this as a method/strategy they had employed organically to bootstrap and move toward broader collaborative opportunities.

Attendees were very enthusiastic about developing a kind of incubator for such projects as an outcome for future meetings of this kind, where short-term projects are launched and attendees collaborate with the goal/design of reporting results at a combined workshop. This was put forward as a possible centerpiece for a follow-on meeting.

Options for Specialized Training

Training and mentoring in close collaboration and partnership across fields is a crucial marker of progress — barriers being overcome — and was demonstrated by some of the earliest successful digital humanities projects, such as Tasman working with Busa in the 1950s[25]: “I personally worked directly with the padre. I had to teach him, in effect, the IBM machines, and at the same time he was teaching me linguistics.” (*IBM engineer Paul Tasman, speaking about his work with Father Roberto Busa in the 1950s using IBM accounting machines and punch cards to develop a concordance of religious texts.*)

Attendees agreed that there are now emerging examples of a training ecosystem and that the chance of success for broader impact improves with shared perspectives and a robust set of options[26]. These programs should be encouraged and connected to

[22] https://artsandculture.google.com/partner/life-photo-collection
[23] https://pro.europeana.eu/page/common-european-data-space-for-cultural-heritage
[24] Watrall E. Building Scholars and Communities of Practice in Digital Heritage and Archaeology. *Advances in Archaeological Practice*. 2019;7(2):140-151. P. 142. doi:10.1017/aap.2019.1
[25] https://www.ibm.com/history/dead-sea-scrolls
[26] Watrall E. Building Scholars and Communities of Practice in Digital Heritage and Archaeology. *Advances in Archaeological Practice*. 2019;7(2):140-151. doi:10.1017/aap.2019.1

one another to the degree possible to continue to build coherence around themes that are common threads. Some examples of training and mentoring programs now available, which can serve as a kind of blueprint, include:

- Heritage Science Fellows Programs and Digital Cultural Heritage Fieldschools ;
- Getty Scholars-in-Residence programs;
- Courses offered at Sorbonne Université around the Plémo 3D platform, open to Sorbonne University PhD students, Master's students, art history and archaeology research professionals and museum staff. Covers topics such as photogrammetry and lasergrammetry.

In order to capitalize upon existing opportunities and solutions for overcoming barriers, and to build a continued strong case for investment in advanced training options, workshop participants suggested some key mechanisms that can facilitate the collaborative work of AI and cultural heritage:

- Training that is intentionally designed to bridge the scholarly divide between two disparate academic communities.
- Short-term cross-training examples, where, for example, an AI developer shadows a heritage professional for short periods and vice versa, collecting and exchanging information and making the “lessons learned” from that experience broadly available.
- Establishing “ambassadorships,” where short-term exchanges across disciplines can encourage researchers sitting alongside each other to observe, answer questions, and collect valuable cross-disciplinary information.

We identified the five major themes (bespoke funding; improved tools; open data sets; small-scale rapid development projects; and training options) as avenues for durable advancement of AI/HS work. There were also other ideas at the workshop — some very specific and strategic, others broadly defined — that were “all over the map.” One clear point is that collective conversation can elicit extremely powerful ideas for how to move forward, and these ideas are often based on real-world experience and professionally-informed intuition from experts. The “solution” of resources/investment is always at the very top of the list, recognizing that investment is the key that unlocks other creative ways to meet challenges and overcome barriers. Workshop attendees were extremely confident that increased investment would produce unprecedented broader impacts, addressed in the next section.

6. Broader Impacts

The workshop participants were clear in their belief that breakthroughs in heritage science can lead to deep and meaningful broader impacts. As one participant put it,

> "These technologies give us deep insight into humanity. We are able to understand things about ourselves that we couldn't otherwise without this technology. This is really a moment of humility for humanity. We should allow computers to teach us, take what we can learn, and bring new questions to the surface as well as ways to answer them." (Workshop Participant)

Considering the myriad of research opportunities and directions, why should time, energy, and resources be devoted to the field of heritage science and, in particular, to the application of AI to cultural heritage studies? The consensus answer lies in the fact that cultural heritage can serve as a premiere domain in which solutions to a vast array of humanity's challenges can be developed. The list below describes specific features of heritage science where broader impacts are directly possible.

Encouraging Social Change

AI & HS initiatives can foster mature intercultural understanding, helping to address societal challenges around diversity and inclusion. In its Convention for the Safeguarding of the Intangible Cultural Heritage, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) notes that an understanding of the heritage of various people encourages inclusivity, intercultural dialogue, and mutual respect for other ways of life[27]. As a portal to a wide variety of communities, including those that are historically underserved, the AI & HS community can play a crucial role in promoting such understanding. The processes and outcomes of collaborative projects will enable connections between cultures and cultivate mutual respect among diverse communities.

Addressing Environmental Concerns

UNESCO also notes that HS can play a pivotal role in environmental studies. "As a living heritage, the body of knowledge, values and practices of intangible cultural heritage related to the environment has the capacity to evolve and adapt for a more sustainable use of natural resources when necessary, permitting communities to better face natural disasters and the challenges of climate change[28]." Recent research also emphasizes that "An understanding of the causes and consequences of previous climate changes is essential for testing present-day climate models and projections" and calls for robust use of the 'paleoenvironmental archives' that archaeological sites provide[29]."

---

[27] UNESCO. (n.d.). "What is Intangible Cultural Heritage?" UNESCO-Intangible Cultural Heritage. Retrieved 8/12/2024, from https://ich.unesco.org/en/what-is-intangible-heritage-*00003*

[28] UNESCO. (2015). Intangible Cultural Heritage and Sustainable Development. UNESCO. https://www.unesco-ichcap.org/wp-content/uploads/2017/08/34299-EN.pdf

[29] St. Amand, F., Childs, S. T., Reitz, E. J., Heller, S., Newsom, B., Rick, T. C., Sandweiss, D. H., & Wheeler,R. (2020). Leveraging legacy archaeological collections as proxies for climate and environmental research. Proceedings of the National Academy of Sciences, 117(15), 8287-8294.

Increasing STEM Access

A mature and vibrant AI and HS community can attract students interested in such areas as art, literature, and archaeology, for example, to the STEM fields germane to heritage science, like chemistry, materials science, and computer science. Research from the United States National Academy of Science, Engineering and Medicine points out that one's own cultural heritage offers an effective access point to STEM. "To achieve equity, practitioners must consider ways to connect the home and community cultures of diverse groups to the culture of science[30]." For example, calls to make scientific education more meaningful to Native American populations by connecting it more directly to their lives and indigenous cultures can progress by offering a platform facilitating engagement with science through the lens of cultural heritage.

In fact, melding the culture and structure of heritage and humanities studies with that of STEM should be a key outcome of these initiatives. The National Academies of Sciences, Engineering, and Medicine recently suggested that it is indeed "the current culture and structure of STEMM [sic] that systematically disadvantage members of these underrepresented groups relative to white and Asian-American men[31]." For example, women – particularly those of color – continue to be underrepresented in STEM fields relative to their presence in the workforce and the U.S. population[32]. Women make up less than 19% of PhDs in Computer Science and Physics, 22% of PhDs in Engineering, less than 30% of PhDs in Math and Astronomy, and less than 40% in Earth Sciences and Chemistry. In contrast, fields key to the study of natural and cultural heritage have a strong female presence. More than 77% of PhDs in Art History are awarded to women, as are approximately 60% of those in Literature and Anthropology and more than 50% in Archaeology[33]. By creating and promoting opportunities that combine STEM disciplines, in which women may be interested but reluctant to pursue because stereotypes lead them to believe they cannot succeed, with heritage-oriented fields where successful women are well represented, we can build upon initiatives already in place to recruit females and other minority students interested in STEM but fleeing its halls.

Developing Future Researchers

An established AI & HS community will provide a transdisciplinary nexus of expertise, allowing faculty, staff, and students from across the humanities and the sciences to

---

[30] National Research Council 2010. Surrounded by Science: Learning Science in Informal Environments. Chapter 7: Culture, Diversity, and Equity. Washington, DC: The National Academies Press. https://doi.org/10.17226/12614).

[31] St. Amand, F., Childs, S. T., Reitz, E. J., Heller, S., Newsom, B., Rick, T. C., Sandweiss, D. H., & Wheeler, R. (2020). Leveraging legacy archaeological collections as proxies for climate and environmental research. Proceedings of the National Academy of Sciences, 117(15), 8287-8294.

[32] St. Amand et al. 2020.

[33] Meyer, M., Cimpian, A., & Leslie, S.-J. (2015). Women are underrepresented in fields where success is believed to require brilliance. 6(235). https://doi:10.3389/fpsyg.2015.00235. .

come together to build, install, and test state-of-the-art HS instrumentation and software tools. Students will gain experience by analyzing HS challenges, investigating solutions, and innovating over infrastructure and tools needed to apply those solutions. Example programs could include REUs, internships, training for research and commercial constituents, certificate programs for education and training, as well as new courses in various disciplines.

Mentoring and apprenticeship opportunities will further aid recruitment of underrepresented groups to STEM. Minority students in STEM programs are more likely than their non-minority counterparts to switch to non-STEM majors[34]; but high quality interactions with faculty and authentic lab experiences, like those these initiatives will offer, can retain them. “These authentic laboratory experiences socialize students into the norms and practices of STEM research, helping them develop science or engineering identity that improves their likelihood of persistence in STEM degree programs...Students who participate in these experiences are also more likely to persist to degree completion and pursue graduate study or careers in STEM[35].”

Fostering Interdisciplinary Collaborations

As new technology-based approaches continue to emerge and be adopted across all disciplines, there is a pressing need for vocabularies, funding structures, and research infrastructures that integrate the humanities/cultural and natural heritage disciplines with those of computer and other sciences. The importance of bridging scholarly cultures should not be understated as a potential broader impact. Indeed, investment in AI and HS represents a unique opportunity to unify a large, diverse set of constituents – students, practitioners, policy-makers, researchers, scientists, and engineers – around the true foundational drivers of the advances that are being made in HS: scientific principles, instrumentation, and large-scale data science.

AI and HS initiatives will enable interdisciplinary convergence around the “Big Data” of natural and cultural heritage collections and environments, where collaboration between computer scientists and heritage scholars is crucial. For example, hundreds of thousands of archaeological sites have been documented across the U.S. over the past 100 years as a result of federal, state, and local legislation. Between 1985 and 2013 alone, close to one million artifacts and antiquities were curated in federal, public, or tribal lands[36]. The call has grown in recent years for leveraging this data to address complex contemporary social challenges[37], including pivotal questions on human-environment interactions, migration, and the emergence of social inequalities amid the backdrop of the multiple spatial and temporal scales offered by the

---

[34] Riegle-Crumb, C., King, B., & Irizarry, Y. J. E. R. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. Educational Researcher, 48(3), 133-144.

[35] Hughes, B. E. (2018). Coming out in STEM: Factors affecting retention of sexual minority STEM students. Science Advances, 4(3), eaao6373. doi:10.1126/sciadv.aao6373.

[36] U.S. Department of the Interior’s Secretary Report, National Park Service.

[37] Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. A., Limp, W. F., Maschner, H. D. G., Michener, W. K., Pauketat, T. R., Peregrine, P., Sabloff, J. A., Wilkinson, T. J., Wright, H. T., & Zeder, M. A. (2014). Grand challenges for archaeology. Proceedings of the National Academy of Sciences, 111(3), 879-880. https://doi.org/10.1073/pnas.1324000111

archaeological record[38]. Knowledge gained from what has been termed a "data-driven big picture archaeology"[39] is now considered critical for making informed forecasts and policy decisions.

Translational Science

One clear benefit from collaborative work is the translational opportunity to diffuse technical approaches into new fields. The challenges inherent in the application of AI to cultural heritage – including data that is spatial/temporal, smaller data sets generally, lack of data standardization and agreed ontologies, ethical and access issues regarding ownership and storage of data – offer many opportunities for the advancement of AI and machine learning that will have broad applicability to other fields. In addition, fields like conservation and cultural heritage experience different demographic challenges than those common to computer science and engineering, and the collaboration between these divergent communities can together strengthen and inspire a broader, more diverse participation. In addition, funding challenges can also be better addressed when collaborative researchers share the burden of developing research support across a number of fields rather than within the limited resources of only one or two disciplines.

[38] Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. A., Limp, W. F., Maschner, H. D. G., Michener, W. K., Pauketat, T. R., Peregrine, P., Sabloff, J. A., Wilkinson, T. J., Wright, H. T., & Zeder, M. A. (2014). Grand challenges for archaeology. Proceedings of the National Academy of Sciences, 111(3), 879-880. https://doi.org/10.1073/pnas.1324000111
[39] Anderson, D. G. (2017). Using CRM data for "big picture" research. In F. P. McManamon (Ed.), New perspectives in cultural resource management. Routledge.

7. Appendix

Final workshop program and agenda for meeting 1 (Getty, July 13-16, 2022):

“Fostering Collaborative Breakthroughs in Heritage Science Through

Machine Learning and Data Science”

Experts Meeting: FINAL PROGRAM

Meeting Program Overview

This NSF-funded meeting will assemble a group of experts to discuss the following themes in AI and Heritage Science and produce a report for the NSF:

- **The Current Landscape**: What case studies represent successful, exciting work that shows the promise of ML/DS methods when applied to problems in Heritage Science?
- **Collaborations and Programs**: What programs and partnerships exist—or are needed—to encourage meaningful collaborations? How can we break down barriers to entry and build pathways for the next generation heritage scientist?
- **Infrastructure**: What equipment, tools, processes, and facilities currently exist or need to be created to provide “Next Generation Heritage Science” capabilities? How can we foster innovation and access?
- **Grand Challenges**: What problems or obstacles continue to hinder the study, analysis, and discovery of heritage objects? Which of these areas offer the most promising, high-likelihood potential for AI-inspired breakthroughs in Heritage Science and, conversely, for Heritage Science-driven innovations in AI?

These themes, and the discussions generated by the myriad expertise of the participants, will reveal not only the promises of ML/DS-based approaches to heritage science, but also the barriers preventing their use and diffusion. Key interactions and ideas from these discussions and presentations will inform a white paper on “Heritage Science through Machine Learning and Data Science” that will be delivered by the organizers. The report will call on the broader Heritage Science and AI Research communities to work alongside the managers of tangible heritage (conservators, curators, professional staff in cultural heritage institutions) to accomplish the following goals:

1. Build new programs
2. Build, deploy, and maintain scientific infrastructure
3. Encourage and develop research funding pathways
4. Develop AI-inspired, novel solutions to longstanding challenges
5. Develop new insights and approaches in AI and Data Science from the uniquely challenging and varied data derived from heritage science collections

Meeting Program Details

The meeting will be held at the Getty and at the Luxe Hotel (Brentwood) in Los Angeles, California, on

Wednesday July 13-16, 2022. The FINAL PROGRAM is as follows:

Wednesday July 13: Arrival and Reception **Luxe Sunset Boulevard Hotel: 11461 Sunset Blvd. Los Angeles, CA 90049** **+1 310-476-6571**	
3-5pm	Participants arrive, check-in at the Luxe Sunset Boulevard hotel
5:30pm	Arrival reception: hotel patio (snacks and drinks near the hotel reception area, to the right off the lobby)

Thursday July 14: Meeting Kickoff and Day 1 **Ada Louise Huxtable Lecture Hall, Getty Research Institute**	
8:30am	Continental breakfast available at the Getty (transportation to Getty provided)
9am	Meeting Kickoff: Welcome, Objectives, and Intro Lightning Rounds
10:30am	Coffee
11am	**Moderated Session: "The Current Landscape"** (Moderated by Brent Seales with participation from selected attendees) This opening conversation will focus on current breakthroughs and promising directions. A review of exciting case studies will highlight the breadth of the landscape covered by emerging work.
12:30pm	Lunch
2:00pm	**Breakouts: "The Possibilities of Collaborations and Programs"** Smaller groups will continue the morning's discussions, focusing on the particular need for collaborations and programs to foster emerging and future work.
3:30pm	Coffee
4pm	**Round-up: "Consensus Themes and Ideas"** Participants will reassemble as a group to hear summaries from each breakout discussion. Moderated conversation will follow.
5:30pm	End of session
5:45pm	Reception followed by dinner in the Getty Private Dining Room (Transportation provided after dinner back to hotel)

Friday July 15: Meeting Day 2 **Luxe Sunset Boulevard Hotel**	
8:30am	Continental breakfast available at the hotel
9am	Meeting: Reflections and Responses to Day 1

10:30am	Coffee
11am	**Moderated Session: "Infrastructure and Frameworks for Collaboration"** (Moderated by Tom Learner with participation from selected attendees) Participants will review and discuss the infrastructure needed for emerging work. Various components and topics will be addressed, including imaging methods, computational needs, the heterogeneity of materials, mobile capabilities, scale, contracts, intellectual property, and ethical frameworks, for example.
12:30pm	Lunch
2pm	**Breakouts: "Grand Challenges"** Small groups will continue the morning's discussion regarding infrastructure with the goal of eliciting some of the most exciting and promising "grand challenge" problems in AI/DS and Heritage Science.
3:30pm	Coffee
4pm	**Round-up: "Consensus Themes and Ideas"** Participants will reassemble as a group to hear summaries from each breakout discussion. Moderated conversation will follow.
5:30pm	End of session
	Dinner on your own

Saturday July 16: Meeting Summary and Action Items **Luxe Sunset Boulevard Hotel**	
8:30am	Continental breakfast available at the hotel
9am	**Breakouts: "Conference Themes"** Small groups will distill key themes and concepts from earlier sessions and prioritize them for report inclusion.
10:30am	Coffee
11am	**Round-up: "Future Directions"** Participants reassemble to discuss action items and group recommendations for reporting. A meeting summary and final details for deliverables will be developed.
Noon	Lunch: hotel patio
1:30pm	Departures and free time

Final workshop program and agenda for meeting 2 (June 19-20, 2023, Hosted by the Observatory of Heritage at Sorbonne University (OPUS) and the Sorbonne Center for Artificial Intelligence (SCAI))

Fostering Collaborative Breakthroughs
in Heritage Science Through Machine Learning and Data Science

19-20 JUNE 2023
SCAI, Pierre et Marie Curie Campus of Sorbonne Université

This workshop is a two-day event to explore ways to leverage artificial intelligence and data science to enhance research and conservation efforts in cultural heritage. Co-hosted by the Observatory of heritage (OPUS) and the Sorbonne Center for Artificial Intelligence (SCAI), the event will be held on June 19-20, 2023, on the Pierre et Marie Curie Campus of Sorbonne Université. The conference will feature presentations from various experts in the field, including speakers from Sorbonne Université (SU), CNRS, the University of Kentucky, and various invited guests from an NSF Experts Meeting on the topic that was held at the Getty in July 2022. The conference aims to facilitate discussions and collaborations among researchers, scholars, and professionals working in heritage science, as well as to showcase innovative approaches to studying and preserving cultural heritage using AI and data science techniques.

Monday, June 19th

9:00 AM - Welcome at SCAI (find the map here)

9:30 AM - General Presentation of SU, its Institutes, and Potential Actions
Discussion led by Nathalie Ginoux, Observatory of heritage (OPUS) , with input from Gérard Biau, Sorbonne center of artificial intelligence (SCAI).

—

Participant Presentations
Thematic talks (3 speakers, 20 minutes each, followed by 15 minutes of group discussion)

10 AM - ARCHAEOLOGY – Leveraging Artificial Intelligence in the Exploration of Ancient History through Archaeological Excavation

[confirmed] **Josef Wilczek, Sorbonne Université**
"Discrimination of wheel-thrown pottery surface treatment by Deep Learning"
The study of pottery surface treatment is essential to understanding the techniques used by ancient potters, in order to explore the cultural and economic organization of past societies. Pottery is one of the most abundant materials found in archaeological excavation, yet classification of pottery surface treatments remains challenging. The goal of this study is to propose a workflow to classify pottery surface treatments automatically, based on the extraction of images depicting surface geometry, calculated from 3D models. These images are then classified by Deep learning.

[confirmed] **Stefania Merlo, McDonald Institute**
"What and who for? Reflections on the use of AI in the exploration of the past through

archaeological practice"
The number of research areas explored and papers published on the use of AI (and in particular machine learning) in archaeology are increasing exponentially. To explore but also contextualise this development in the discipline, I first briefly describe how AI is used today in archaeological practice at a landscape and site level. Then, using the case studies of the metsemegologolo - https://www.metsemegologolo.org.za/wordpress/and MAEASaM (Mapping Africa's Endangered Archaeological Sites and Monuments – https://maeasam.org/) projects, both based in and catering for audiences in Africa, I reflect on the practical and ethical issues of the use of AI in low resourced environments.

[confirmed] Corey Toler-Franklin, University of Florida
"Multispectral Analysis and Deep Learning for Heritage Science and Life Science Research"
Many cultural heritage artifacts and biological specimens in natural history collections are more comprehensively understood by multispectral analysis. Infrared imaging permits the study of subsurface materials hidden under pigments, and has been used to examine underdrawings of oil paintings. Ultraviolet fluorescence reveals original color patterns on colorless fossils for species classification. However, faded color, and material layers that exhibit subsurface scattering and spatially varying surface reflectance make it difficult to reconstruct the shape and appearance of natural materials. This talk presents a texture transfer framework that reconstructs invisible (or faded) appearance properties in organic materials with complex color patterns. I will motivate the project with a study that computes surface orientation (normals) at different material layers as a function of emission wavelength for effective scientific analysis in life science. Key contributions include a novel ultraviolet illumination system that records changing material property distributions, and a color reconstruction algorithm that uses spherical harmonics and principles from chemistry and biology to learn relationships between color appearance and material composition and concentration. Finally, I will explain a novel algorithm that extends the effective receptive field of a convolutional neural network to detect and analyze multiscale features of interest in large high-resolution image datasets.

Discussion

—

11:15 AM - TEXT – Analyzing written heritage with natural language processing and artificial intelligence
Speakers:

[confirmed] Louis Falissard, Sorbonne University, Bibliotheque National de France
"Learning when to search"
While Large Language Models (LLM) are able to accumulate and restore knowledge, they are still prone to hallucination. Especially when faced with factual questions, LLM cannot only rely on knowledge stored in parameters to guarantee truthful and correct answers. Augmenting these models with the ability to search on external information sources, such as the web, is a promising approach to ground knowledge to retrieve information. However, searching in a large collection of documents introduces additional computational/time costs. An optimal behavior would be to query external resources only when the LLM is not confident about answers. In this paper, we propose a new LLM able to self-estimate if it is able to answer directly or needs to request an external tool. We investigate a supervised approach by 1) introducing a hallucination masking mechanism in which labels are generated using a close book question-answering task, and 2) leveraging parameter-efficient fine tuning techniques to train our model on a small

amount of data. Our model manages to answer directly to 73.2% of the known answers and prefers to search for 79.3% of the unknown answers.

[confirmed] **Lise Jaillant, Loughborough University-Language processing**
"Unlocking Written Heritage with AI"
To analyse written heritage, researchers and other users first need access to that heritage. Yet, it is extremely complicated to get access to archival records originally created in digital form (such as emails, PDFs, Word documents, and audio-visual files). This lack of access is due to a wide range of reasons – including data protection, sensitivity, copyright and technical difficulties.

For example, the archive of the British writer Ian McEwan contains paper and born-digital records, but only the paper part of the collection is available to researchers at the Harry Ransom Center in Austin, Texas. Literary writers are not the only record creators whose digital archive is inaccessible. Government professionals create born-digital records on a daily basis, and the most important records are then sent to archival repositories to be preserved for prosperity. But without access to these materials, researchers will struggle to write the history of our recent past.

This talk will address the following questions:

- How can we unlock born-digital records using Artificial Intelligence?
- What are the risks of applying AI to archives?
- What is the best way to work across disciplines (including the Humanities and Computer Science) to solve the problem of locked archives?

[confirmed] **Stephen Parsons, University of Kentucky**
"Ink-ID and Trans-modal rendering"
The conventional wisdom has been that carbon-based ink is not visible in micro-CT, making the virtual unwrapping of manuscripts using that kind of ink a useless endeavor. In this talk I will show that techniques in Machine Learning allow for the evidence of carbon-based inks to be detected in micro-CT. This result changes the thinking around virtual unwrapping, and has also opened the door to an associated technique we call "trans-modal rendering." The idea is to image in one modality (micro-CT) and then render in another (full color photography) based on a trained network that transfers the style from one modality to another. I will show current results from material from Herculaneum and the Dead Sea Scroll collection.

Discussion

—

12:30 PM - Lunch Break (On-site)
Discussion over lunch

—

1:30 PM - CONSERVATION – Artificial Intelligence and Object Conservation in Museums or Libraries

[confirmed] **Emeline Pouyet - Tsveta Miteva**
"Artificial Intelligence for hyperspectral imaging of historical paintings"
Hyperspectral imaging (HSI) in the visible and SWIR domains are fast and non-invasive imaging methods that have been adapted by the field of conservation science to study painted surfaces.

By measuring the reflectance at a given pixel on a 2D surface, the resulting 3D hyperspectral data cube contains millions of recorded spectra. While processing such large amounts of spectral data poses an analytical and computational challenge, it also opens new opportunities to apply powerful methods of multivariate analysis for data evaluation. With the intent of expanding current data treatment of hyperspectral datasets, and solving the problem of nonlinear unmixing of hyperspectral reflectance data acquired on painted works of art, innovative data analysis approaches based on the use of AI have been recently developed. The efficiency and limitations of the proposed methods for painted surfaces from cultural heritage will be presented and discussed through the study of laboratory prepared paint mock-ups, and historical paintings.
Keywords: hyperspectral data analysis, data reduction and visualization, non-linear unmixing, painted surfaces

[confirmed] **Vincent Christlein, Friedrich-Alexander-Universität**
"Pose Estimation for Artworks Retrieval"
Human pose detection represents an important factor in scene analysis of artworks. Additionally, pose estimation can be used for an interpretable image retrieval. This talk will discuss pose estimation and its role in image retrieval in various fields of application such as classical archaeology and art history.

[confirmed] **Henry Kautz, University of Rochester.**
"Large Language Models for Heritage Science: Potential and Pitfalls"
In 2023, AI systems employing large language models (LLMs) can hold fluent conversations on practically any topic, write code, extract data from scientific literature, and design original artifacts. We explore ways that the current generation of LLM-based systems can be used in heritage science and describe heritage science tools that would be feasible to build. Along the way we shall describe pitfalls that users may stumble into when using LLMs and ways to avoid them.

Discussion

—

2:45 PM - 3D VISUALIZATION – Data Visualization / Cultural Heritage-The Promise of AI

[confirmed] **Nicolas Leys, Chloé Guennou**
"A taste of real-time scientific visualization : facial reconstructions and Palais-Royal theater reconstitution (Molière, Facial reconstruction)"
3D Scientific visualization provides representation of simulated phenomena and sampled data with applications in application fields. Our team has recently pushed forward scientific visualization applied to digital humanities, using cutting-edge real-time 3D engines.

Face reconstruction aims at recovering the facial appearance of an individual from the sole data of the underlying skull. It is an important problem for forensics, anthropology, archeology or history. Based on the previous work of the Facile algorithm team, we have built a complete process to be able to offer a customization of the reconstructed faces, allowing researchers to offer realistic renderings. In this presentation, we will present our reconstruction of a Napoleonic soldier, so-called Christophe Cros.

The Palais-Royal theater in Paris was considered as one of the most important and luxurious in Paris in its early stages, and hosted Molière and his troupe in the 17th century. Entirely

destroyed by fire, we work together with art historians to virtually reconstruct the theater in 3D. Our reconstruction summarizes the state-of-the-art knowledge about this theater and will be presented in the form of a virtual visit during this presentation.
The objective of this presentation is to present our advances in digital humanities, to discuss possible AI integration, despite AI is not part of the research activities of our team.

[confirmed] Marzia D'Angelo, University of Florence; Centro Internazionale per lo Studio dei Papiri Ercolanesi 'M. Gigante' in Naples (Italy).
"Exploring the Potential of 3D Imaging and AI in Maque-IT: Towards an Enhanced Software for the Virtual Reconstruction of Herculaneum Papyri"
Due to carbonization and extreme compression of the Herculaneum scrolls, their mechanical unrolling often resulted in the presence of stratified layers of papyrus, which compromised the immediate reading of the text. Maque-IT is the first image import and editing software specifically conceived to facilitate the reordering of layers in virtual reconstructions of the Herculaneum papyri. The software enables manual selection of misplaced layers identified under the microscope and automatically repositions them to their original position based on established stratigraphic 'rules' of stratigraphy. This presentation explores the potential of integrating papyrological expertise, 3D imaging, and AI to enhance Maque-IT. The implementation of the software with new 3D images of Herculaneum papyri being developed by EduceLab could hold promising prospects for automatic layer recognition. Compared to the flattened 2D images currently in use, the three-dimensional perspective of the fragments would facilitate the observation of layer overlap, while RGB 3D photos could allow for improved identification of surface breaks through color differences and fiber continuity. The collaboration between Maque-IT and the new software pipeline for revealing hidden texts of the Herculaneum papyri could revolutionize scroll reconstruction, allowing for the movement of both visible and previously concealed layers of text.

[confirmed] Seth Parker, University of Kentucky
"Photogrammetry + enhancement engine using ML/AI"
The world-class collection of Herculaneum papyri (held at four separate institutions) features carbonized papyrus in various states of conservation. For papyrus that came from carbonized scrolls, we see a large number of opened trays with visible text. In this talk I will discuss a system and a method for acquiring aligned spectral images and a photogrammetric reconstruction of thousands of trays for the purpose of better modeling of that material - and furthering an enhancement approach never before possible, based on ML/AI. Results from Naples and Oxford have established a baseline we will also apply at the Institut de France's library and at the British Library.

Discussion

4:00 PM - Break

4:30 PM - Metadata, Applications, Tools for Transformation, Vision

[confirmed] Christy Chapman, University of Kentucky
"XAI for Heritage Science"
Advanced imaging techniques – 3D modeling, spectral photography, and volumetric x-ray, for example – can be applied to all types of cultural objects and can be combined to create complex digital representations comprising many disparate parts. Emergent technologies like virtual unwrapping and artificial intelligence (AI) make it possible to create "born digital" versions of unseen features, such as text and brush strokes, that are "hidden" by damage and therefore

lack verifiable analog counterparts. Thus, the need for transparent metadata that describes and depicts the set of algorithmic steps and file combinations used to create such complicated digital representations is crucial. This talk discusses our work with Metadata Encoding Transmission Standard (METS) to meet emerging digital provenance metadata needs.

[confirmed] **Federica Nicolardi, Università degli Studi di Napoli Federico II; Centro Internazionale per lo Studio dei Papiri Ercolanesi 'Marcello Gigante'**
"Reading and Reconstructing the Scrolls from Herculaneum: New Perspectives and Challenges of a Three-dimensional Approach"
The Herculaneum scrolls' extremely precarious state of conservation, charred and fragmented into many thousands of pieces as they are, represents the major obstacle to reading them and to the systematic study of their texts. In most cases, reading a Herculaneum papyrus cannot be an immediate activity, but rather turns out to be a 'mediated' task. Virtual reconstruction is an instrument for mediating between the poorly preserved scrolls and their texts. Tremendous progress has been made in the last twenty years thanks to the application of more and more advanced photographic techniques. However, much can still be done to refine our methodologies and techniques for reconstructing the Herculaneum scrolls, in particular by enhancing and enriching the long work of scholars with a new digital approach. This talk will discuss potentialities of three-dimensionality applied to virtual reconstruction. The possible benefits of a 'reverse' approach – from virtual unwrapping to virtual re-wrapping– will be addressed and discussed.

[confirmed] **Emma-Jane Alexander, Gary Quasebarth, MechDyne**
"Tripping the Digital Fantastic (The Landscape of Experiential Storytelling)"
Mechdyne brings 26 years' experience in using visualization to underpin the storytelling aspects of our teaching, research, and entrepreneurial ecosystems. Persistent digital transformations are redefining the roles of and connections between the real world, and virtual worlds. Mechdyne would like to present brief case study examples of where we have supported research and educational communities of practice. Challenges and opportunities that these new spaces and tools bring will be addressed, and our appetite for play, failure and relentless innovation in the realm of visualization and associated AI technologies. Emma-Jane and Gary will deliver a creative free flowing presentation and discussion session based around their combined experiences in higher education and industry.

Discussion

6:00 PM - 9:00 PM: Apéro jussieu, Le patio 1526

Apero sponsored by Mechdyne

(Dinner on your own)

Tuesday, June 20th

Morning: Venue is the Institut Nationale d'Histoire des Arts (INHA)

9:30 AM Presentation on Plemo 3D at the André Chastel Center (Sorbonne Université) (2nd floor - room 232)

10:15 AM Plenary Session- Salle Perrot : "Promising Technologies, Collections, Discoveries"
Group Discussion led by Dr. Henry Kautz, Christy Chapman, and Frédérique Andry-Cazin

12:00 PM Lunch at Le Mesturet (near Centre Chastel)

(Travel to the Datalab at the Bibliothèque Nationale de France)

2:30 PM - 4:00 PM Tour, Demos, Discussion
HANDWRITING ANALYSIS at the Bibliotheque National de France Datalab and discussion on Handwritten Textual Recognition (Confirmed)

4:00 PM Wrap Up: Findings, Recommendations, Themes
Summary findings for reporting and expert notes/recommendations led by Dr. Brent Seales

5:15 Adjourn