Plan Overview

A Data Management Plan created using DMPonline

Title: INFO4730-AI-Driven-Cultural-Heritage-and-Repatriation-Repository-AICHRR

Creator: Joey Richardson

Principal Investigator: joey richardson

Affiliation: Other

Template: DCC Template

Project abstract:

AI and Cultural Heritage: Digital Repatriation of Artifacts Project Purpose

The purpose of this project is to collect and archive metadata related to cultural heritage artifacts to analyze trends in digital repatriation efforts. This repository specifically focuses on how AI tools such as image recognition and provenance tracking assist in identifying artifact origins and automating legal documentation.

Metadata Sources

The data in this repository was collected and structured from the following cultural heritage archives:

- **UNESCO Digital Heritage Database**
- **Smithsonian Open Access**
- **British Museum Collection**

Dataset Fields

The metadata.csv file includes the following standardized fields:

1. **Artifact Name**: The common name of the object.
2. **Origin**: Country or region of origin.
3. **Date of Creation**: Estimated time of manufacture.
4. **Description**: Brief physical or historical context.
5. **Current Location**: The museum or archive currently holding the item.
6. **Ownership History**: Known provenance and chain of custody.
7. **Repatriation Status**: Current state of return (Requested, In Progress, or Completed).
8. **AI Usage**: The specific AI technology applied to the repatriation case (Image Recognition).

Repository Structure

- /Global-Repatriation-Dataset: Contains the full collection of 5 artifacts.
- /Exhibits/British-Museum-Claims: A sub-folder containing artifacts specifically located in the British Museum to analyze institutional patterns.

ID: 202147

Start date: 13-04-2026

Last modified: 14-04-2026

INFO4730-AI-Driven-Cultural-Heritage-and-Repatriation-Repository-AICHRR

Data Collection

What data will you collect or create?

I will collect core descriptive metadata, including the names, origins, and physical descriptions of five high-profile artifacts sourced from institutions like the British Museum and UNESCO. Additionally, I will create specialized fields to track provenance history and repatriation status, such as whether a return is "In Progress" or "Requested." Most importantly, I will generate new data regarding AI interventions by documenting the specific types of tools like image recognition or database matching, used to verify artifact origins and support legal claims. This structured dataset will be stored in a non-proprietary CSV format to ensure long-term accessibility and transparency for researchers and legal advocates.

How will the data be collected or created?

I will collect artifact metadata by extracting and restructuring existing records from open-access museum databases and international cultural heritage archives. These narrative descriptions will be repackaged into standardized fields including origin, current location, and provenance history to ensure data consistency. Additionally, I will create new metadata by analyzing case studies to identify specific AI applications, such as image recognition or database matching, used in repatriation efforts. This combined dataset will be organized into a multi-level CSV structure to facilitate institutional analysis and long-term research access.

Documentation and Metadata

What documentation and metadata will accompany the data?

To ensure the data is reusable and professional, I will provide a **README.txt** file containing the project rationale, field definitions, and source origins for the user group. Each artifact will be accompanied by a standardized metadata schema that includes core descriptors like origin and current location, as well as specialized legal and technological tracking fields. Furthermore, I will include administrative metadata within the file headers to maintain data integrity and document the evolution of repatriation claims. This documentation will be stored within the root directory of the archive to provide a complete and transparent context for all users.

Ethics and Legal Compliance

How will you manage any ethical issues?

To manage ethical issues, I will ensure that all data is sourced from open-access repositories and credited to the original institutions to respect intellectual property and data ownership. I will implement ethical provenance tracking by providing historical context for looted items, ensuring that the narratives of displaced cultural groups are represented accurately and respectfully rather than just as legal statistics. Furthermore, the repository will be maintained with transparency regarding AI biases, including notes on the limitations of the algorithms used to identify or verify these artifacts. All metadata will be reviewed to avoid offensive or culturally insensitive terminology, aligning the archive with international ethical standards for heritage documentation.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

To manage copyright and Intellectual Property Rights (IPR) issues, I will only ingest metadata from open-access databases and public domain sources that explicitly allow for academic or non-commercial reuse. All entries in the repository will include a specific field for original source attribution to ensure that the primary institutions, such as the Smithsonian or UNESCO, receive proper credit for their data. I will apply a Creative Commons Attribution license to the restructured dataset, which ensures the metadata remains an open resource while protecting the project's intellectual contribution. Furthermore, I will avoid the

inclusion of copyrighted high-resolution imagery or restricted legal documents, focusing instead on text-based metadata that describes these items within the fair use framework for educational research.

Storage and Backup

How will the data be stored and backed up during the research?

During the research, the data will be stored on a secure, password-protected cloud drive to ensure it is accessible for editing while remaining protected from unauthorized access. To prevent data loss, I will maintain a primary working directory and a secondary backup folder on a separate physical or cloud-based storage system. I will also implement a manual version control system by saving dated copies of the CSV files at the end of each work session. This redundant storage strategy ensures that a clean and accurate version of the metadata is always available even if a file becomes corrupted or accidentally deleted.

How will you manage access and security?

I will manage access by utilizing a private, password-protected cloud storage platform that restricts entry to authorized team members only. To enhance security, I will implement two-factor authentication on all accounts linked to the repository to prevent unauthorized login attempts. Furthermore, the final dissemination of the data will involve a transition to a read-only public directory, ensuring that the general user group can view the metadata without the ability to edit or delete the files. This approach balances the need for collaborative research security with the project goal of open information provision for the public.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

The data of long-term value includes the **standardized metadata records** for the selected artifacts, as these provide a permanent digital snapshot of provenance and repatriation status at a specific point in history. The **AI intervention metadata** is also critical for preservation, as it documents the evolving role of technology in cultural heritage and serves as a reference for future researchers investigating algorithmic accuracy. Additionally, the **README documentation** and the associated **metadata schema** must be retained to ensure that future users can interpret the data structure and understand the project's original context. These components will be preserved in a non-proprietary CSV format to guarantee that the information remains accessible and machine-readable for long-term legal and academic use.

What is the long-term preservation plan for the dataset?

To ensure long-term preservation, the dataset will be converted into stable, non-proprietary formats such as **CSV** for data and **PDF/A** for documentation to prevent software obsolescence. The final archive will be deposited into a dedicated digital repository or an institutional data warehouse that provides permanent identifiers and ensures bit-level preservation. This plan includes a commitment to periodic integrity checks and a metadata strategy that complies with international archival standards to keep the information discoverable for future researchers. This approach ensures that the data remains accessible and usable even as specific hardware and software technologies evolve over time.

Data Sharing

How will you share the data?

I will share the data by uploading the complete archive to a secure, shared digital folder and providing a direct access link to the designated user group. To ensure maximum accessibility, the repository will be disseminated via a read-only public directory that allows researchers and legal advocates to view and download the CSV files without risking the integrity of the master records. I will also include the **README.txt** and metadata schema within the same directory to provide the necessary context

for immediate use. This method allows for wide-reaching distribution across various institutional platforms while maintaining a centralized version for consistent referencing.

Are any restrictions on data sharing required?

There are no major legal restrictions on sharing the metadata, as all core information is derived from open-access museum databases and public domain sources. However, I will implement a selective restriction on the "AI intervention" notes if they contain sensitive details about active, ongoing legal disputes that could compromise a repatriation claim. To manage this, any data deemed sensitive will be anonymized or summarized at a high level to protect the strategies of the claimant groups while still providing research value. This approach ensures that the project remains as open as possible while respecting the confidentiality and safety of the cultural heritage advocates involved.

Responsibilities and Resources

Who will be responsible for data management?

As the lead researcher, I will be primarily responsible for the day-to-day management, quality control, and security of the dataset throughout the project lifecycle. This responsibility includes overseeing the accurate extraction of metadata, maintaining the folder hierarchy, and ensuring that regular backups are performed to prevent data loss. I will also be accountable for final data preservation and ensuring that the repository is successfully transitioned to its long-term storage platform at the conclusion of the research. For collaborative phases, I will define clear access permissions to ensure that any additional contributors adhere to the established metadata standards and ethical guidelines.

What resources will you require to deliver your plan?

To deliver this plan, I will require access to **open-access digital archives** and museum databases, such as those provided by the British Museum, UNESCO, and the Smithsonian, to extract primary metadata. I will also need a **secure cloud storage platform** with sufficient capacity for redundant backups and a folder-based version control system to manage the CSV files. Additionally, I will utilize **standard spreadsheet and text editing software** to restructure the data and author the project documentation. Finally, I will require **reliable internet connectivity** to facilitate data collection and the eventual dissemination of the repository to the targeted user groups.

Planned Research Outputs

Image - "Benin Bronze Head"
https://digitalbenin.org/items/204

Image - "Parthenon Sculptures"
https://www.britishmuseum.org/collection/object/G_1816-0610-1

Model representation - "Saint Tirumankai Alvar"
https://jameelcentre.ashmolean.org/object/EA1967.165

Model representation - "Bust of Nefertiti"
https://id.smb.museum/object/606573

Model representation - "Rosetta Stone"
https://www.britishmuseum.org/collection/object/Y_EA24

Publication - "Museum of Looted Antiquities"
Review: "The Museum of Looted Antiquities (MoLA) is a virtual gallery dedicated to the intricacies of tracing the smuggling, looting, purchase, and eventual return of antiquities repatriated since 1950. With contributions from community members–academics, lawyers, museum professionals–MoLA presents a comprehensive collection of trafficked antiquities with over 100 publicly available entries. Created to consolidate documentation tracing an artifact's journey from its looting to eventual repatriation, MoLA aspires to preserve all records of ownership and exchange of an object for future research."

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Benin Bronze Head		Image	Unspecified	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No
Parthenon Sculptures		Image	Unspecified	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No
Saint Tirumankai Alvar		Model representation	Unspecified	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No
Bust of Nefertiti		Model representation	Unspecified	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No
Rosetta Stone		Model representation	Unspecified	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No
Museum of Looted Antiquities	10.17613/87qxg-82r44 ...	Publication	2025-01-01	Open	tDAR		None specified	LIDO (Lightweight Information Describing Objects)	No	No