Data Curation Domain Profile
GIS Applications and Spatial Data Practices
in Cultural Heritage Preservation

IS262A Data Management & Practice

Rosalia Iriye
Jack Morrissey
Catherine Whalen

Introduction

Scholars have recognized the important role of information professionals in facilitating research data management in the digital age.[1] With this in mind, the Fall 2025 “Data Management and Practice” class at University of California, Los Angeles, was assigned a final project inspired by Purdue University’s Data Curation Profiles project.[2] This meta assignment called for our group to discover and analyze data about data management practices, culminating in a Data Curation Profile (DCP) for a specific field and subfield of our choosing. Throughout the quarter, our lessons and readings focused on researcher best practices in the field of Information Science, such as the procedures outlined in the Open Archival Information System (OAIS) Reference Model for data archiving, Findable Accessible Interoperable and Reusable (FAIR) Data Principles aimed at enhancing data reusability, and Transparency, Responsibility, User focus, Sustainability and Technology (TRUST) Principles for digital preservation of data in reliable repositories.[3] These concepts played key roles in our own decision-making processes and in our analysis of our findings on the data management choices made by others.

Project Proposal

Our group proposed a Data Curation Domain Profile for the field of Cultural Heritage Preservation with a subfield of Geographic Information Systems (GIS) Applications and Spatial Data Practices. Our goal was to deliver a dataset of one hundred open access articles published

---

[1] Lisa Federer, “Research Data Management in the Age of Big Data: Roles and Opportunities for Librarians,” *Information Services & Use*, *36* (2016): 35. https://doi.org/10.3233/isu-160797; Carol Tenopir et al., “The Time Has Come…To Talk About Why Research Data Management Isn’t Easy,” *Proceedings of Charleston Library Conference* (2019): 298–301. https://dx.doi.org/10.5703/1288284317185.
[2] “Data Curation Profiles Directory,” Purdue University Libraries, accessed November 15, 2025, https://docs.lib.purdue.edu/dcp/.
[3] Consultive Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS), Recommended Practice, Issue 3*: CCDS 650.0-M-3 (2024); Mark Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship” (2016). https://www.nature.com/articles/sdata201618; Dawei Lin et al., The TRUST Principles for digital repositories” (2020). https://doi.org/10.1038/s41597-020-0486-7.

between 2020 and 2025, accompanied by a DCP, and to deposit our work in an appropriate repository. The articles we selected encompass cultural heritage preservation studies and incorporate GIS applications or other spatial data research methodologies. Our article dataset coding template included the following headers (metadata): Contributor, DOI/URL, Title, Author, Journal, Publisher, Publication Year, Data Collected, Data Deposited, DOI/URL of Data Deposit, Data Repository, DOIs of Data Paper, Funding Source, Country of Study, Province/Municipality, and Software Used. The attached README contains a Data Dictionary and additional information regarding article collection, data entry and data cleaning and standardization protocols that were employed throughout the project.[4]

Journals and Funding

After an initial search of repositories via JSTOR, ResearchGate, Semantic Scholar and University of California Libraries, our group identified nine journals from four publishers that had generated a sufficient number of articles that met the criteria outlined in our project proposal: *Built Heritage* (SpringerOpen), *Digital Applications in Archaeology and Cultural Heritage* (Elsevier ScienceDirect), *Heritage Science* (Springer Nature Nature Partner Journal), *ISPRS International Journal of Geo-Information* (MDPI), *Journal of Cultural Heritage* (Elsevier ScienceDirect), *Land* (MDPI), *Remote Sensing* (MDPI), *Scientific Reports* (Springer Nature), and *Sustainability* (MDPI). Just over half of the articles in the dataset came from MDPI (52%), followed by Springer Nature (26%), Elsevier ScienceDirect (18%) and SpringerOpen (4%).[5]

[4] Iriye, R., Morrissey, J., & Whalen, C. (2025) Data Curation Domain Profile: GIS Applications and Spatial Data Practices in Cultural Heritage Preservation, v1.0 [Dataset]. OSF. https://doi.org/10.17605/OSF.IO/X32U5.
[5] At the time of dataset creation and analysis, the relationship of SpringerOpen to Springer Nature was in transition (https://support.springernature.com/en/support/solutions/articles/6000281876-springer-nature-brand-websites-are-moving-to-springer-nature-link) and policies of the SpringerOpen journal differed from those of the Springer Nature journals. As such, project contributors decided to treat these as separate publishers for the purposes of the project.

Interestingly, we found little overlap in funding for the entire dataset, with the exception of nine studies that were either fully or partially funded by the National Natural Science Foundation of China. This minimal overlap indicated that sponsorship of relevant research may be more ad hoc in the realm of cultural heritage preservation studies incorporating GIS applications and spatial data practices than in some other specializations where financial backing originates from a limited number of sources. Twenty-three articles stipulated that they had no external funding and five did not mention funding at all. Much of the financial backing for the remaining seventy-two articles was tied to governments, universities or special projects, primarily on a geographical basis.

Applicable Metadata Standards: Cultural Heritage Preservation

The methodological inclusion of field- and subfield-specific metadata is critical to ensuring that research datasets are discoverable and (re)usable.[6] Observing geospatial elements in cultural heritage preservation reveals how interoperability expands metadata standards and data management software development across our field and subfield. International Organization for Standardization ISO21127 for cultural heritage information was created with the adoption of formal ontology CIDOC Conceptual Reference Model (CRM). The most recent ISO edition cites Open Geospatial Consortium (OGC) standards, adding “4 (sub)classes and 17 properties to align with OCG standards for geospatial data [...]”.[7] CIDOC further supports the bridging of our field and subfield. For example, it promotes extensions to OGC standards, such as CRMgeo for

---

[6] Chao, T.C. (2014), “Enhancing Metadata for Research Methods in Data Curation.” Proc. Am. Soc. Info. Sci. Tech., 51: 1-4. https://doi.org/10.1002/meet.2014.14505101103.

[7] ISO 21127:2023, *Information and documentation — A reference ontology for the interchange of cultural heritage information*, (Geneva: International Organization for Standardization), accessed December 10, 2025, https://www.iso.org/standard/85100.html.

querying geospatial data,[8] and facilitates the ongoing development of the International Core Data Standard for Archaeological and Architectural Heritage.[9]

Building on these emerging global data standards, Getty data management platform Arches offers open-source software for geospatial and cultural heritage preservation.[10] Arches integrates conceptual models (CIDOC; Linked Art; content standards by geographic area) and controlled vocabularies (Getty Art and Architecture Thesaurus; Forum for Information Standards in Heritage [FISH]). The software's description outlines its interoperability with commercial and open formats (JSON, geoJSON) across both fields, and adherence to World Wide Web Consortium (WC3) standards for data modeling.[11] Targeted metadata standards for GIS applications in cultural resource management, such as The United States Federal Geographic Data Committee (FDGC)-sponsored National Park Service (NPS) Cultural Resource Spatial Data Transfer Standards, continue to develop in regional contexts.[12]

In defining cultural property,[13] United Nations Educational, Scientific and Cultural Organization (UNESCO) outlines two types of tangible heritage: movable (artifacts that can be transported) and immovable (conservation sites; buildings).[14] While it is not a metadata standard, this movable/immovable distinction aligns with UNESCO's geospatial resources, from their

[8] CIDOC CRMgeo, "CRMgeo Home," CIDOC, accessed December 10, 2025, https://cidoc-crm.org/crmgeo.
[9]International Core Data Standard for Archaeological and Architectural Heritage (Working Draft), accessed December 10, 2025, https://www.cidoc-data.org/aswg-international-core-data-standard-for-archaeological-and-architectural-heritage.
[10]Arches Resource Model Working Group (ARM WG), "Arches Modeling Documentation," accessed December 10, 2025, https://www.archesproject.org/arm-wg-documentation.
[11] "Standards and Interoperability," *Arches Project*, accessed December 9, 2025, https://www.archesproject.org/standards-and-interoperability.
[12]"Cultural Resources GIS Data Standards," *National Park Service*, accessed December 10, 2025, https://www.nps.gov/subjects/historicpreservationfund/cultural-resources-gis-data-standards.htm.
[13] "Convention for the Protection of Cultural Property in the Event of Armed Conflict with Regulations for the Execution of the Convention," *UNESCO*, accessed December 9, 2025, https://www.unesco.org/en/legal-affairs/convention-protection-cultural-property-event-armed-conflict-regulations-execution-convention.
[14] Ibid.

geodatabase World Heritage Sites Online Map Platform (WHOMP), hosted through Esri ArcGIS, to UNESCO's commitment to open spatial data through GitHub availability and QGIS software plug-ins.[15]

Applicable Metadata Standards: GIS Applications and Spatial Data Practices

To fully understand the nature of geospatial data deposits in open access journals, it is necessary to examine how open standards for geospatial data are shaped and implemented for open-source and commercial GIS software. Sponsored by the FDGC, Content Standard for Digital Geospatial Metadata (CSDGM) provides a national metadata standard for geospatial data.[16] In contrast to non-geospatial metadata standards, CSDGM is notable for its mandate concerning how geospatial data is constructed, with multiple sub-elements for describing research methods such as "entity" and "processor," which can denote software types.[17] Building on FDGC CSDGM, ISO XML schema ISO 19139 and respective content standard ISO 19115 allow for global compatibility when encoding geographic metadata.[18] National Archives and Records Administration (NARA) guidelines for linked open database formats are inclusive of proprietary data types, particularly those associated with American GIS software company Esri.[19]

Software ArcMap[20] and its more commonly known successor, ArcGIS,[21] use a default metadata style called "Item Description," which simplistically provides "Description" and "Properties"

[15] "Geospatial information for peace," *UNESCO*, accessed December 9, 2025, https://www.unesco.org/en/geospatial.
[16] "Standards | Content Standard for Digital Geospatial Metadata (CSDGM)" Federal Geographic Data Committee, n.d. https://www.fgdc.gov/metadata/csdgm-standard.
[17] Chao, T.C. (2014), "Enhancing Metadata for Research Methods in Data Curation." Proc. Am. Soc. Info. Sci. Tech., 51: 1-4. https://doi.org/10.1002/meet.2014.14505101103.
[18] "ISO/TS 19139-1:2019 Geographic information — XML schema implementation" International Organization for Standardization, 2019. https://www.iso.org/standard/67253.html.
[19] "Linked Open Data for Database Formats" National Archives and Records Administration (NARA), September 11, 2025. https://www.archives.gov/preservation/digital-preservation/linked-data/databases.
[20] "Metadata Styles and Standards - ArcMap | Documentation." ArcGIS Desktop, n.d. https://desktop.arcgis.com/en/arcmap/latest/manage-data/metadata/metadata-standards-and-styles.htm.
[21] "View Metadata - ArcGIS Pro | Documentation." ArcGIS Pro, n.d. https://pro.arcgis.com/en/pro-app/latest/help/metadata/view-and-edit-metadata.htm.

tabs with additional fields (File Path; Title; Summary; Description; Author; Credits; Tags; Hyperlink base; Last Saved; Last Printed; Last Exported; Default Geodatabase) alongside elements integral to the software's interface, such as thumbnails. Other metadata standards can be selected within the platform, including "FDGC Content Standard for Digital Geospatial Metadata metadata content, all ISO 19139 metadata content, all North American Profile metadata content, and all INSPIRE metadata content."[22] Regardless of how metadata is imported, metadata for all objects is available in XML document format.[23] Autodesk Revit supports FGDC CSDGM Standard and ISO 19139.[24] QGIS does not enforce a metadata standard, however, its internal schema is DublinCore compatible,[25] with plug-ins such as GeoCat Bridge,[26] which can be used to import and export metadata with various ISO standards.

Collected Data Types

Our group chose a field and subfield that encompassed both humanities and sciences, which we believed would yield an interesting mix of quantitative and qualitative data. Our research affirmed this prediction, producing a diverse set of data types. Due to the wide range in both types and descriptors used by researchers, we conducted cleanup of the data types associated with each of our 100 articles to ensure that we would accurately capture overlap in our analysis.

---

[22]"The ArcGIS Metadata Format - ArcMap | Documentation." ArcGIS Desktop, n.d. https://desktop.arcgis.com/en/arcmap/latest/manage-data/metadata/the-arcgis-metadata-format.htm.
[23] Ibid.
[24]"About Metadata." Autodesk Map 3D 2024, 2024. https://help.autodesk.com/view/MAP/2024/ENU/?guid=GUID-AA374714-4990-4CC5-B73B-1E997784C37E.
[25] "Class: QGIS Layer Metadata" QGIS Python API, n.d. https://qgis.org/pyqgis/3.40/core/QgsLayerMetadata.html.
[26]"Metadata Editing" GeoCat Bridge for QGIS, n.d. https://docs.geocat.net/bridge/qgis/latest/metadata_editing.html#:~:text=GeoCat%20Bridge%20provides%20a%20basic,Metadata%20language.

Geospatial data can be divided into two data types: vector data and raster data. Vector reflects discrete values as geometric objects, whereas raster is continuous on a cell grid.[27] We utilized the following acronyms to denote and standardize sources for these data types: Close Range Photogrammetry (CRP) photogrammetric data; Computer-Aided Design (CAD) data; Digital Elevation Model (DEM) data; Digital Surface Model (DSM) data; Digital Terrain Model (DTM) data; Electrical Resistivity Tomography (ERT) data; Ground Penetrating Radar (GPR) data; Light Detection and Ranging (LiDAR); Synthetic Aperture Radar (SAR); Terrestrial Laser Scanner (TLS) data; Uncrewed Aerial System (UAS) photogrammetric data and Unmanned Aerial Vehicle (UAV) photogrammetric data.

Shapefiles, developed and maintained by Esri, are the most common forms of geospatial vector data, comprising three file formats: .shp for points, lines and polygons attributes such as geographic coordinates, roads and buildings, respectively, .shx for indexing .shp and .dbf table for non-geometric attributes.[28] Raster images like photogrammetry and satellite imagery can be processed to create 3D models, with data types and respective formats such as Digital Elevation Models (DEM) (commonly formatted as TIFF; GeoTIFF; IMG; Esri GRID; SID), Digital Surface Models (DSM) in Computer-Aided Design (CAD) formats and Digital Terrain Models (DTM), which can reflect any of the prior model formats dependent on ground features.[29] Light Detection and Ranging (LiDAR) point clouds are vector data types with common formats such

[27] Raúl Jiménez Ortega. “What’s Special about Geospatial Data?” ArcGIS Blog, Esri, August 27, 2025. https://www.esri.com/arcgis-blog/products/developers/developers/whats-special-about-geospatial-data.
[28]“Shapefile file extensions” ArcGIS Desktop, Esri, n.d. https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/shapefile-file-extensions.htm.
[29] Rakesh Malhotra. “Elevation Data Formats” John A. Dutton Institute for Teaching and Learning Excellence, College of Earth and Mineral Sciences, The Pennsylvania State University Penn State, 2009. https://www.e-education.psu.edu/geog480/node/513.

as ASCII, Esri SHP or LAS; point clouds can be further processed into the aforementioned models.[30]

Despite a fair amount of overlap in data types used in GIS applications and spatial data practices, our dataset showed a wide variety of ways in which researchers refer to and utilize both subfield-specific data and historical data used for the study of cultural heritage preservation. In addition to the geospatial data types outlined previously, we encountered numerous kinds of quantitative and qualitative data drawn from tangible and intangible cultural heritage. Examples include sound propagation in rock art landscapes, urban planning and water use records for ancient cities, folklore accounts, historic sports records and present-day visitorship data for cultural sites. The varying degrees to which these resources were coupled with geospatial methods yielded an expansive "Data Collected" column with four hundred unique instances of data types. This portion of our dataset proved most difficult to standardize.

Our contributors relied on Exploratory Data Analysis to view and organize the "Data Collected" column of the coding template. First, we generated a word cloud, which was used to identify the most prevalent instances of field- and subfield-specific data. The word cloud also helped us locate and remove redundancies or nondescript terms we had initially employed to describe data contained in the dataset. Due to the diversity of cultures studied, we decided to retain most data concerning cultural heritage in its original form, as it was referred to in each article. This was done in an attempt to remain conscious of the ways scalability can obscure the value of diversity in different global research settings and to "validate local experiences of data without removing them from a more global network of information exchange."[31]

---

[30] Ibid.

[31] Katie Rawson and Trevor Muñoz, "Against Cleaning," in *Debates in the Digital Humanities 2019*, ed. Matthew K. Gold and Lauren F. Klein (University of Minnesota Press, 2019), 279–92.

Addressing additional inconsistencies, the term "historical" was used to replace the word "historic" wherever it appeared, as "historical" was favored by the vast majority of researchers. Further, the term "institutional" was initially incorporated as an antecedent to data from overly specific public and private sources. Where possible, data of this nature was ultimately substituted with more useful descriptive nomenclature. The terms “photos” and “orthophotos” were replaced with “photography” and “orthophotography,” respectively. Finally, instances of the most prevalent software, ArcGIS and QGIS, that provided version numbers were clearly delineated using the term "version" before the version number, and those without versions were appended with "version unspecified."

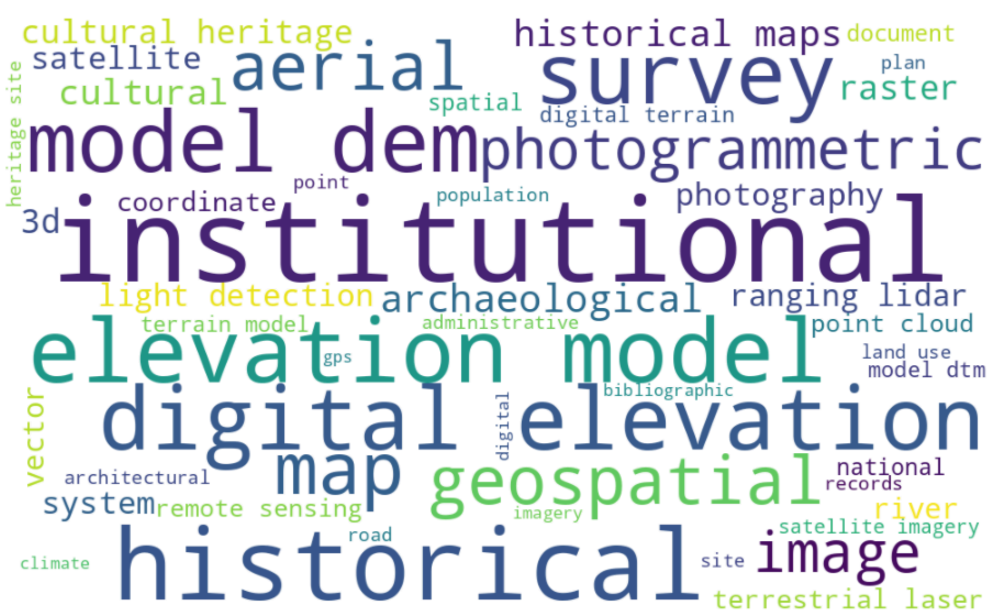

Figure 1: Data type word cloud before cleaning, generated using Python WordCloud library.

Figure 2: Data type word cloud after cleaning, generated using Python WordCloud library.

Other Findings

After encountering a significant number of global cultural heritage preservation studies early in the article sourcing process, our group expanded the coding template to include a column denoting the geographical locations where research within each article was conducted. This "Country of Study" column revealed that the research we gathered took place in a total of thirty-six countries, with the highest percentage of studies coming from China (37%), followed by Spain (12%) and Italy (7%). Due to the high frequency of studies in China, and to provide a more complete picture of GIS applications and spatial data use in cultural heritage preservation research, another column for specific provinces/municipalities associated with research performed in China was also added. Of note, two of the one hundred studies were not associated with single geographic locations. One was a global study encompassing nine hundred ninety-two UNESCO World Heritage Sites; the other was a museum-based application of research originating from a single Spanish institution's collection database that incorporated cultural heritage resources from forty-one countries.

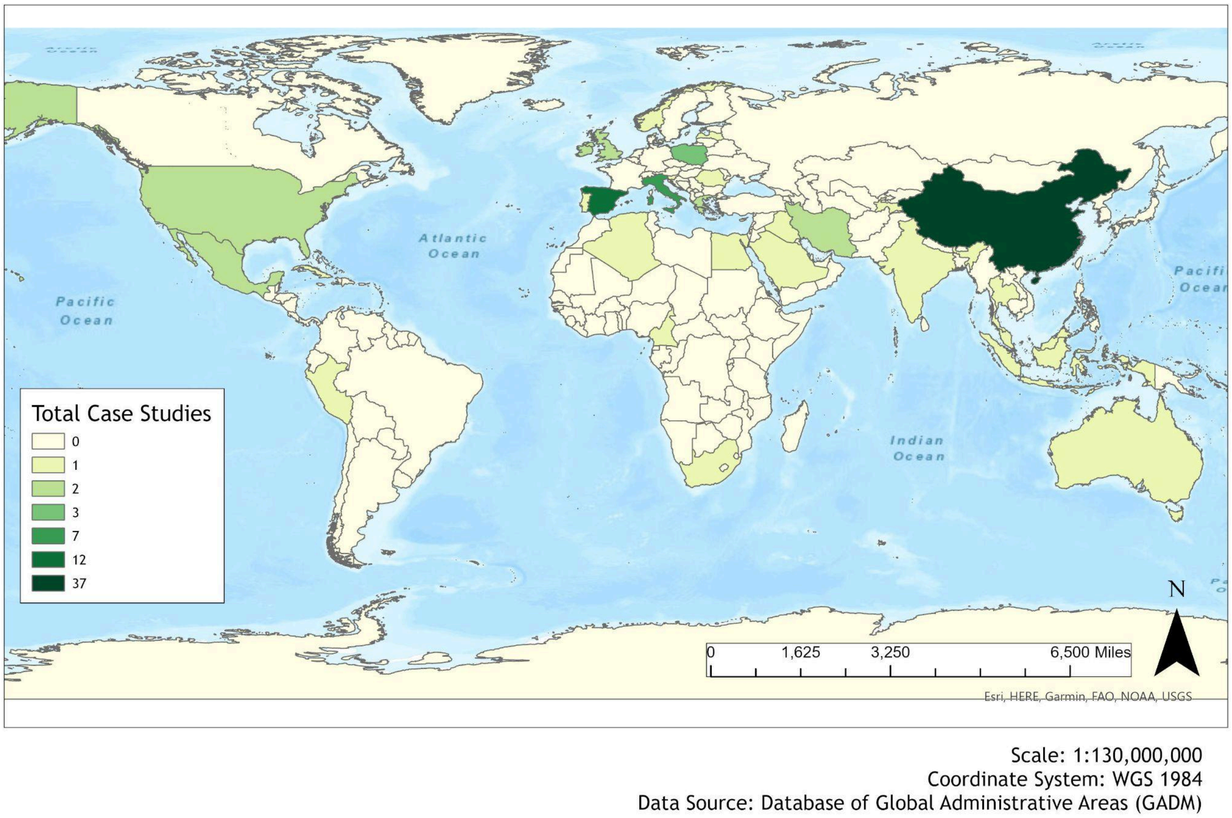

Figure 3: Mapping of total journal article case studies by countries based on data analysis using Pandas library and visualization in ArcGIS.

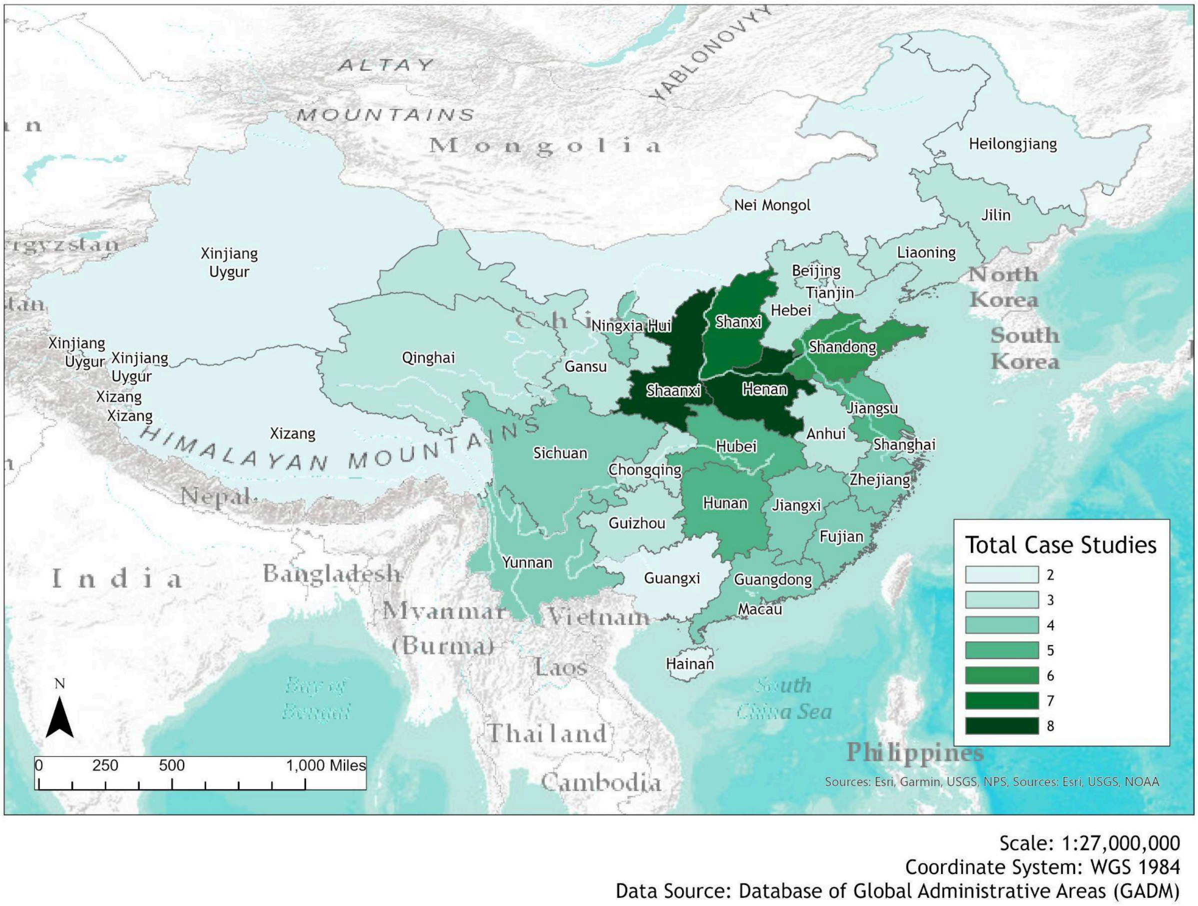

Figure 4: Mapping of total journal article case studies in Chinese provinces and autonomous regions based on data analysis using Pandas and GeoPandas libraries and visualization in ArcGIS.

Many of the articles in our dataset examined the use of geospatial software in cultural heritage research. To capture this additional data point, we further expanded the coding template to include a column indicating which software was used in each study. Of the one hundred articles, fifty-two incorporated the use of more than one kind of software. Overall, we encountered one hundred sixty-six unique instances of GIS and GIS-related software. Once again, we found that the ways in which researchers referred to or utilized a specific software were non-standardized, particularly if a version was included. It is clear, however, that ArcGIS/ArcGISPro and QGIS were the most commonly employed software types, with sixty-four and twenty-three instances of each, respectively. Beyond version numbers, instances of ArcGIS software use were inconsistently denoted. Most authors simply indicated use of "ArcGIS," but some referenced earlier versions ArcMap/ArcGIS Desktop, while others named successor ArcGIS Pro. Version 10.8 was the most widely used iteration of ArcGIS and all twelve instances originated from research based in China.

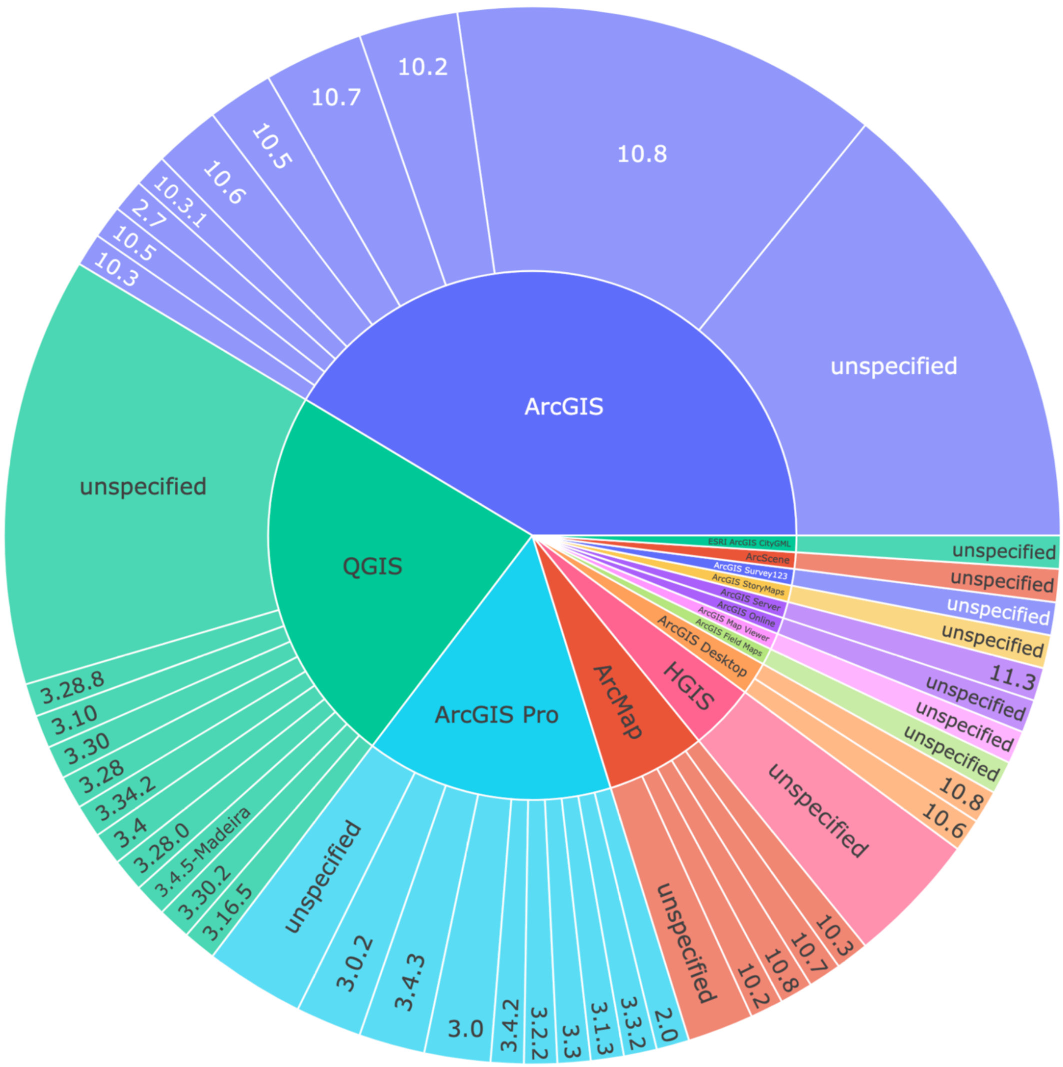

Figure 5: Sunburst plot of software version differentiations among ArcGIS and ArcGIS derivatives, QGIS and HGIS, generated using Plotly Python library.

Policies and Requirements:

Data Availability Statements, Repositories and Reproducibility

Information professionals have observed an uptick in requirements from journals and funders.[32] In a 2018 case study of scientific journal PLOS ONE, for instance, Lisa Federer et al. investigate author compliance with data sharing policies through an analysis of Data Availability Statements (DAS). For the purposes of our Data Curation Profile, we applied the Data Availability Statement coding categories shared by Federer et al. to our analysis of DAS when populating the "Data Deposited" column of our dataset.[33] Due to the variety of funding sources reflected in the dataset, it was not possible to investigate individual sponsor policies within the timeframe of the project, however, our group was able to identify the requirements of journals and publishers that generated the articles. The following breakdown of requirements and findings from our dataset analysis yields valuable information regarding the effect of such requirements on compliance.

MDPI publications *International Journal of Geo-Information*, *Land*, *Remote Sensing* and *Sustainability* provide detailed submission instructions for authors, including MDPI's Research Data Policies regarding supplementary materials, data deposit and source code.[34] These policies are informed by TOP Guidelines, promoting best practice in openness and transparency in research.[35] Notably, MDPI requires DAS for all articles and provides recommended statements for various scenarios. MDPI also strongly urges authors to preserve data in trusted repositories

---

[32] Carol Tenopir et al., "Data Sharing, Management, Use, and Reuse: Practices and Perceptions of Scientists Worldwide," *PLoS ONE*, 15(3), e0229003 (2020). https://doi.org/10.1371/journal.pone.0229003.
[33] Lisa Federer et al., "Data Sharing in PLOS ONE: An Analysis of Data Availability Statements," *PLoS ONE*, 13(5): e0194768 (2018). https://doi.org/10.1371/journal.pone.0194768.
[34] "International Journal of Geo-Information: Instructions for Authors," *MDPI*, accessed November 16, 2025, https://www.mdpi.com/journal/ijgi/instructions; "Land: Instructions for Authors," *MDPI*, accessed December 7, 2025, https://www.mdpi.com/journal/land/instructions; "Remote Sensing: Instructions for Authors," *MDPI*, accessed November 26, 2025, https://www.mdpi.com/journal/remotesensing/instructions; "Sustainability: Instructions for Authors," *MDPI*, accessed November 16, 2025, https://www.mdpi.com/journal/sustainability/instructions.
[35] "TOP Guidelines," *Center for Open Science*, accessed November 16. 2025, https://www.cos.io/initiatives/top-guidelines.

for a minimum of five years post-publication. Formal data citation is encouraged, as is the release of any code, software information and supplemental data used in studies. Reflecting MDPI's policies, DAS were incorporated into the vast majority of the publisher's articles, even if they simply indicated "N/A." However, despite the MDPI requirements, five of the fifty-two articles associated with the publisher lacked formal DAS.

SpringerOpen's *Built Heritage*, Springer Nature's *Scientific Reports* and Springer Nature's Nature Partner Journals' *Heritage Science* similarly require DAS or, in the case of *Built Heritage*, statements regarding "Availability of Data and Materials." However, both Nature Portfolio journals go further in requiring availability of not only data but of unique materials associated with it, and that computer code and algorithm used must be made available upon request under separate heading "Code Availability." Their standards also mandate the deposit of certain data types and require various reporting summaries by field and/or material type in addition to DAS. This includes reporting full provenance information for geological, archaeological, and palaeontological research.[36] All *Built Heritage* and *Scientific Reports* articles in our dataset included DAS and only one of the sixteen *Heritage Science* articles in our dataset lacked a Data Availability Statement. One *Heritage Science* article and one *Scientific Reports* article also included Code Availability sections.

Elsevier ScienceDirect's *Digital Applications in Archaeology and Cultural Heritage* and *Journal of Cultural Heritage* take a less stringent approach. The journal encourages – but does not

[36] "Nature Partner Journals Heritage Science: Reporting Standards and Availability of Data, Materials, Code and Protocols," *Nature*, accessed November 26, 2025, https://www.nature.com/npjheritagesci/for-authors-and-referees/about/editorial-policies/reporting-standards; "Scientific Reports: Editing and Publishing Policies," *Nature*, accessed December 7, 2025, https://www.nature.com/srep/journal-policies/editorial-policies#availability; "SpringerOpen Editorial Policies: Availability of Data and Materials," *SpringerOpen*, accessed December 7, 2025, https://www.springeropen.com/get-published/editorial-policies#availability+of+data+and+materials.

mandate – similar policies, including those regarding DAS.[37] Predictably, none of the eighteen Elsevier Science Direct articles included DAS, but one provided a link to their deposited data via GitHub.

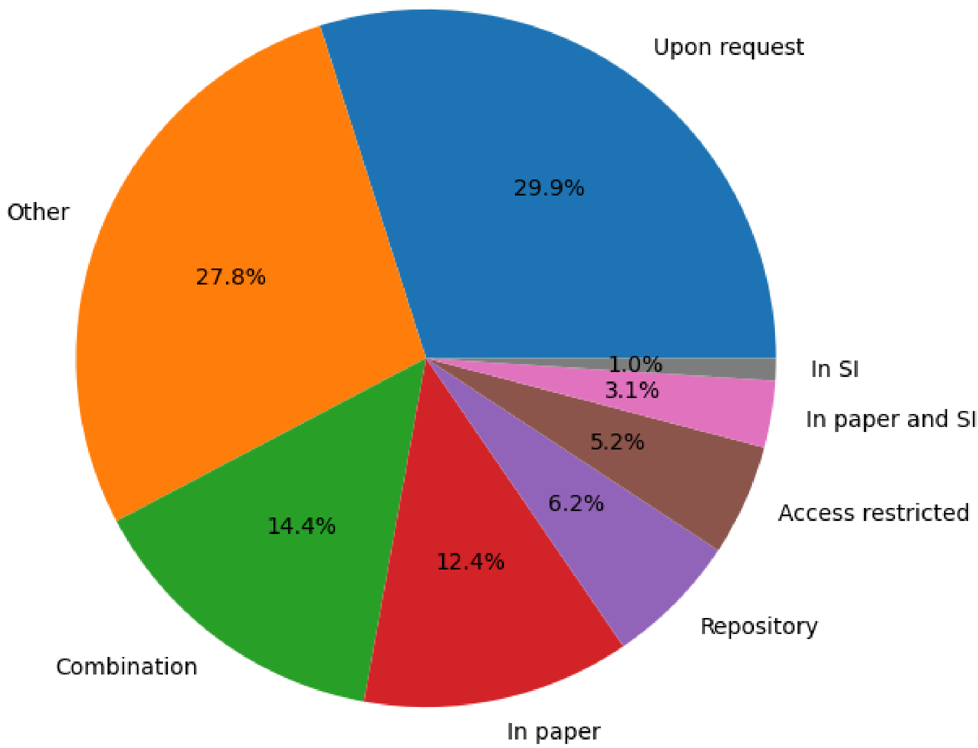

Figure 6: Data availability for our dataset referencing Federer et al. "Data Deposited" coding categories, generated using Matplotlib Python library.

Authors of only thirteen of the one hundred articles in our dataset deposited data in a total of twenty repositories. Data associated with five of the fourteen articles was deposited in two

[37] "Digital Applications in Archaeology and Cultural Heritage: Guide for Authors," *Elsevier ScienceDirect*, accessed December 7, 2025, https://www.sciencedirect.com/journal/digital-applications-in-archaeology-and-cultural-heritage/publish/guide-for-authors; "Journal of Cultural Heritage: Guide for Authors," *Elsevier ScienceDirect*, accessed November 16, 2025, https://www.sciencedirect.com/journal/journal-of-cultural-heritage/publish/guide-for-authors.

repositories; data tied to the remaining articles was placed in just one repository for each. There were a total of four deposits in GitHub, the largest amount of any repository. MDPI, Museo Nacional y Centro de Investigación de Altamira and Zenodo had two deposits each. Additional repositories used included other university-affiliated or government-created databases, e.g. Scotland's Canmore, as well as proprietary repositories like Amazon Elastic Block Store and MongoDB. Of note: the Canmore database is now defunct and redirects to Trove, a new government-affiliated repository[38]. Upon further investigation, the dataset contained within Trove was available only by request. A similar lack of accessibility was common across many of the other data deposits and supplementary materials comprising our dataset. Most data was in unusable and non-raw formats, or was associated with broken links, making reproducibility of the majority of studies extremely difficult, if not impossible.

The Digital Preservation Coalition (DPC) provides guidance for creators depositing geospatial data, as well as digital preservation organizations. It notes the complexities surrounding interoperability between commercial and open formats when addressing various metadata standards and access to proprietary GIS softwares.[39] A shapefile, for example, is proprietary but operates within an open standard. In facilitating access, the organization recommends using open-source software QGIS broadly due to its ability to open both proprietary and non-proprietary formats.[40] The DPC guidelines build upon Library of Congress' recommendations for the "[m]ost complete data (all layers, appendices), even if proprietary, with a preference for preserving the native format and projection of the data," generally pointing to shapefiles, Esri File GeoDatabase, OGC GeoPackage, and GeoJSON formats.

---

[38]Historic Environment Scotland, "Retiral of HES Web Services," *Historic Environment Scotland*, May 12, 2025, last modified June 30, 2025. https://www.historicenvironment.scot/about-us/news/retiral-of-hes-web-services/. historicenvironment.scot.
[39] "Preserving GIS" Digital Preservation Coalition, July 2021. http://doi.org/10.7207/twgn21-16.
[40] Ibid.

Using DPC-recommended National Archives open-source tool DROID (Digital Record Object Identification) for file format analysis of our articles' publicly available datasets yielded 2272 total files. These only contained thirteen shapefiles, ninety-five GeoTIFFs and nothing in the Esri File GeoDatabase, OGC GeoPackage and GeoJSON categories. Shapefiles generally lacked the required three formats, with an uneven ratio of .shp to .shx files. While there are workflows for locating or reconstructing missing .shx in shapefiles such as QGIS Geometry Checker[41] or ArcGIS Pro Repair Geometry[42] tools, incompleteness of geospatial formats impacts data reproducibility.

Key Takeaways

Our study produced a robust dataset of articles reflecting the expected combination of quantitative and qualitative data types. Imaging and historical data were the most frequently used data types. Of these, the top five were Digital Elevation Models (DEM) (twenty-nine instances), historical maps (seventeen instances), Light Detection and Ranging (LiDAR) (ten instances), Digital Terrain Models (DTM) (seven instances) and satellite imagery (seven instances). Geographically, China-based studies were most common, comprising 37% of the dataset, with Shaanxi and Henan as the top provinces. The most used software types were ArcGIS and QGIS.

In creating the dataset, we successfully pulled one hundred articles from nine journals spread over four of the top publishers in the field and investigated their policies and requirements with regard to Data Availability Statements, repositories and reproducibility. While we identified some inconsistencies of note, our analysis showed that more stringent requirements from journals resulted in better compliance. Still, we encountered many articles that lacked DAS or

[41] "25.2.2. Geometry Checker Plugin" QGIS Documentation, QGIS project, October 31, 2025. https://docs.qgis.org/3.40/en/docs/user_manual/plugins/core_plugins/plugins_geometry_checker.html.
[42] "Repair Geometry (Data Management)" ArcGIS Pro, Esri, n.d. https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/repair-geometry.htm.

stipulated that data was only available "upon request." We found relatively few links to data deposits/known repositories and even fewer that worked or linked yielded data in a (re)usable condition. We also discovered that the studies we chose had little overlap in sources of funding, indicating that this particular field and subfield may not rely on a narrow group of financial sponsors for research and rendering a similar investigation into sponsor policies and requirements impractical.

Academics from the scientific community, among others, have noted the significance of responsible data management throughout the research process as a preventive measure against potential loss.[43] Our participation as data producers and analyzers in this project highlighted this importance for us. We directly observed how researcher best practices such as those outlined in the FAIR Data Principles and TRUST Principles - or lack thereof - impact concepts like findability, accessibility, interoperability, reusability, transparency, responsibility, sustainability and technology for digital preservation of data.

We hope that our dataset and analysis will prove useful to researchers who are interested in exploring how GIS applications and spatial data practices can be used in the field of cultural heritage preservation. In the interest of openly sharing our research, we have uploaded our dataset (.csv), README (.txt) and this DCP (PDF) to trusted repositories GitHub under Repository https://github.com/CatherineWhalen/CHP_GIS_SpatialData_DCDP and Open Science Foundation under Project DOI 10.17605/OSF.IO/X32U5, along with the Python code we used (.ipynb project file when exported from Colab) and GeoTIFF maps (.tif), in the noted, universally accessible formats under a CC BY 4.0 license. Pandas, GeoPandas, Matplotlib, Plotly

[43] Robert Downs and Robert Chen, "Chapter 12: Curation of Scientific Data at Risk of Loss: Data Rescue and Dissemination," in *Curating Research Data Volume One*, ed. Lisa R. Johnston (Association of College and Research Libraries, 2017), 271.

and WordCloud libraries were utilized for data analysis and visualization. Our contributors welcome comments and questions. Contact information is provided below.

Rosalia Iriye iriyerosalia@gmail.com
Jack Morrissey jackmorrissey@ucla.edu
Catherine Whalen catherinewhalen@ucla.edu