e-ISSN 2724-3923

magazén
Vol. 6 – Num. 2 – December 2025

A Data Atlas Method for Analysing and Visualising Dispersed Cultural Heritage Collections

Andreas Vlachidis
University College London (UCL), United Kingdom

Isobel MacDonald
The British Museum, United Kingdom

Foteini Valeonti
University College London (UCL), United Kingdom

Julianne Nyhan
TU-Darmstadt, Deutschland

Kim Sloan
The British Museum, United Kingdom

Abstract The history of collecting shows how people and institutions have tried to interpret and shape their worlds, but tracing objects across heritage collections is hindered by fragmented, uneven data. Issues of scope, digitisation, legacy metadata and dispersed systems limit data-driven research and digital reconnection efforts. This paper proposes the Collection Data Atlas, a framework for mapping complex informational landscapes across institutions and epochs. Using the AHRC-funded Sloane Lab, it applies the Sloane Data Atlas to visualise and synthesise the intricate data environment of Sir Hans Sloane's collection.

Keywords Collections as Data. Cultural Heritage. Data Atlas. Sloane. Digital Humanities.

Edizioni Ca'Foscari

Peer review

Submitted 2025-08-18
Accepted 2025-10-30
Published 2025-12-23

Open access

Citation Vlachidis, A.; MacDonald, I.; Valeonti, F.; Nyhan, J.; Sloan, K. (2025). "A Data Atlas Method for Analysing and Visualising Dispersed Cultural Heritage Collections". *magazén*, 6(2), [1-30], 149-178.

DOI 10.30687/mag/2724-3923/2025/02/002

1 Introduction

The proliferation of digital collection databases built by cultural heritage institutions over the past six decades, and the many digitisation projects that have been undertaken in the heritage sector in recent years, have acted to dramatically expand the empirical basis upon which research on the history of collections can be pursued (Institute of Museum and Library Services 2018; Jones 2022, 2). Such developments and digital advances, open the theoretical possibilities to pursue data-driven, long durée analysis of how processes of fluidity have shaped the heritage collections at our disposal today. Understanding the information environment of such collections, which encompasses both historic, handwritten, manuscript-based descriptions of acquired objects as well as present-day digital databases is imperative for examining how cultural heritage collections have grown over time. This can help us examining the development of collections beyond the lens of acquisitions made by individuals or institutions that "captures only one side of collections' history" (Cornish, Driver 2020, 327). Moreover, better understandings of these information environments have the potential to cast new light on the fluidity of collections, the role that circulation played in their formation and mobilisation, and to signify the histories of objects, their acquisition, location and movement over time.

The information that describes heritage objects and collections tends to be highly complex, often having been catalogued by many people, over a long period in a range of analogue and digital formats. It is often siloed across multiple institutions and different institutional recording systems, with varying levels of interoperability and accessibility. This can lead to a lack of understanding of what material is held within a given institution. At present, the ability to undertake such research is reliant on an understanding of the idiosyncrasies of collection databases and related material located across different institutions. Research is moving beyond the accumulation of collections, seeking a deeper examination of the dynamic movement and organisation of objects across information systems, and individual and institutional agents and actors (Driver et al. 2021, 8). Thus, our research aims to address the following questions: How can the review and mapping of dispersed datasets be undertaken in a systematic way, using a methodological framework that can capture shared understandings of complex data landscapes, that have been created by multidisciplinary teams, institutions, and systems with varying data structures? Additionally, how can the data landscape of such heritage collections, including their intersections and the interactions of their constituent parts be synthesised and visualised in a structured and transparent way?

The collection of Sir Hans Sloane is a paradigmatic example of potential that may be unlocked by answering the above questions. Assembled from the 1680s onwards, and in part financed by profits from the transatlantic trade in enslaved human beings, Sloane's vast collection of natural history, pharmaceutical specimens, books, manuscripts, prints, drawings, coins, and antiquities from across the world was made as Britain became a global trading and imperial power (MacGregor 1994, 53). Today Sloane's physical collections, and their historical collection records, are spread across three national institutions: the National History Museum (NHM), British Library (BL) and BM. They are, in turn, recorded in five different digital cataloguing systems across those institutions, each shaped by the individual disciplines and institutional histories that gave rise to them. The *Sloane Lab: Looking back to build future shared collections* (hereafter Sloane Lab) project, funded by the AHRC's *Towards a National Collection Discovery* programme, re-establishes the broken links between Sloane's catalogues and collections across the Natural History Museum (UK), the British Library, and the British Museum, devising automated and augmented methods that are relevant not only for Sloane's collection but also for cultural heritage collections in museums, galleries, libraries, and archives (Nyhan et al. 2025). The project positions Sloane's collection as a microcosm through which to investigate the technical, infrastructural, conceptual, historical and social challenges faced in bringing together digital cultural heritage collections.

A crucial precursor of researching Sloane's collection is the mapping of its informational landscape, both historical and present day, to move from a fractured, institution-specific view of the individual parts of the collection to a view of the collection as a whole, allowing its cross-institutional landscape to be understood in its entirety for the first time. We accordingly present the *Collection Data Atlas* as an instrument that can further this aim, while likewise functioning as a methodological instrument that can be used to identify, represent and ultimately support the mobilisation of complex and cross-institutional collections (see e.g., Institute of Museum and Library Services 2018; Padilla et al. 2019; Towards a National Collection).[1] The focus is not at the lower level of data characteristics and metadata that can be easily addressed by schemas and conceptual models of unified data representations. Instead, the Collection Data Atlas addresses softer silos, which relate to the scope, size, availability, coverage, legacy attributes, and manifestation of collection data, which often persist both within and between institutions. Such attributes cannot be adequately addressed by conceptual data models and metadata

1 See Towards a National Collection: `https://www.nationalcollection.org.uk/`.

mappings alone. Our approach offers a holistic metaphor to allow for a multidisciplinary understanding, inventorying analysis, and conception of collection landscapes.

This paper begins with a synthesis of the wider research context of this paper, including a brief discussion of research that uses data-driven methods to understand the history of collections. An overview of the way in which digital collections data has been used in various projects, alongside the move for transparency in the creation of datasets for analysis, situates the *Collection Data Atlas* as an intentional instrument designed for the early stages of large-scale data aggregation. The Sloane Lab Data Atlas then exemplifies the design rationale and application of the *Collection Data Atlas* in the data aggregation process, demonstrating how it aids data auditing, data ingestion prioritisation and decision-making within the complex data landscape of the Sloane collection. We conclude with a critical reflection on the usefulness of proposed methodological framework within the Sloane Lab and, crucially, its transferability to other collections as data projects. The reflection addresses technical and broader efforts to understand collection histories and presents how other projects can effectively adopt the proposed framework to chart dispersed collection landscapes within the cultural heritage domain and beyond. The Sloane Lab Data Atlas is available on GitHub[2] and a snapshot of it is appended **[Appendix]**.

2 Research Context

2.1 Collections as Data

In recent years, cultural heritage collections have become key sites of technological encounter. As Thomas Padilla has argued, "Collections as data entails thinking about ways to increase meaning making capacity by making collections more amenable to use across an expanded set of methods and tools, typically but not exclusively computational in nature" (Padilla 2017, 2). Efforts to support computationally driven research of cultural heritage collections can cultivate open documentation and content-sharing approaches, with useful examples including the HathiTrust Research Centre, the National Library of the Netherlands Data Services and APIs, the Library of Congress' Chronicling America, and the British Library (15). However, it is worth noting that although Collections as Data projects foreground the practice, theory, and ethics surrounding the use of digital collections data by cultural heritage institutions, many

2 Access the resource online: `https://github.com/sloanelab-org/data-atlas`.

smaller institutions have not built digital collections or designed access to them with the aim of supporting computational use (10).

Meanwhile, data-driven projects tend to use digital collections data for provenance research. The Provenance Lab at Leuphana University, Lüneburg, for example, pursues provenance research through a network-based approach, highlighting social and economic trends in collection formation and change with digital methods (Rother, Mariani, Koss 2023). The *Digital Benin* project, hosted at Markk Museum am Rothenbaum, Hamburg, has created an online knowledge forum, reconnecting object data from 131 institutions worldwide of over 5,000 objects that were looted in the British military campaign against Benin City in 1897 (Luther 2024). With a similar focus on colonial heritage held in museums, the *Pressing Matter: Ownership, Value and the Question of Colonial Heritage in Museums* hosted at Vrije Universiteit in Amsterdam, is creating a Knowledge Graph to aggregate entities across institutional silos (Shoilee 2022). The *CUDAN Open Lab*, based out of Tallin University, is using collection data from twelve European contemporary art museums to explore the acquisition of contemporary art works and associated museum profiles at an aggregate level (Solà et al. 2023). The *Between Canon and Coincidence*, hosted at Leiden University, is studying the provenance of Latin American collections in Europe through the collection and analysis of collection data from 12 museums in 9 countries using network analysis, big data computer models and AI (Leiden University 2023; Berger 2023). These examples highlight the growing use of digital collections data and data-driven methods to augment and challenge our understanding of object provenance and broader patterns of collection growth and dispersal.

In the UK a growing number of research projects use data-driven approaches to examine the history of institutional collections. These have ranged from an examination of relational patterns between the museum and communities associated with the collection of the Pitts Rivers Museum, Oxford (Gosden, Larson, Petch 2007; Petch 2006; 2002); an analysis of objects from a particular part of the world within one museum collection (Wingfield 2011); an analysis of specific areas of a national collection (Phillipson 2019); and the mapping of certain collector typologies (Penn, Cafferty, Carine 2019). A two-year research project at the British Museum conducted a comprehensive quantitative examination of the history of its current collection records (MacDonald 2023). Its aim was to trace patterns across time, material and departments and to provide an institutional context within which individual objects, collections and collectors could be understood. Whilst most of these projects focus on the digital collections data held by one cultural heritage institution, some have traced patterns running across institutions. For example, Charlotte Dixon analysed

data across 13 UK museums to compare the collecting practices of Indian Ocean boat models by British institutions (Dixon 2023, 51).

As this wider research context suggests, within the cultural heritage sector there is an increasing emphasis on virtual reunification and the interlinking and analysis of data from dispersed collections (Punzalan 2014; Hyvönen 2012). These efforts are often led by cross-institutional, large-scale aggregators, such as Europeana (de Boer V et al. 2012). The key stages of a data integration workflow tend to be described as comprising the following main steps: Harvesting, Ingestion, Mapping, Indexing, Storing, Monitoring, Enriching, Displaying, and Publishing (Siqueira, Martins 2022). Yet, the decision-making process for selecting data to be included in (and excluded from) projects has not been documented clearly, with internal dynamics and priorities often defining the scope of data ingestion. Importantly, this is interdisciplinary question that needs curators and technical teams to work on together something that is not always possible or straightforward given the university-museum partnerships, and associated funding schemes it requires.

Data-driven projects are as strong as the data they examine, and within cultural heritage institutions these datasets hold various nuances, biases and limitations. Digital collection databases may be incomplete if part of a collection is uncatalogued. An example of this can be seen in Sloane's collection of coins and medals, catalogued in "Coins vol 1" to "Coins vol 10" and "Medals". They were presumed to be among the few remaining manuscript items in the BM Coins and Medals department when it was hit by an incendiary bomb on the night of 10 May 1941, and they have not been located since the war. These catalogues were described in 1933 as "ten bound foolscap volumes, with indexes in several volumes, […] bound in ten volumes according to countries, apparently all coins" (MacGregor 1994, 164). As this suggests, the indices may have been included in the ten volumes and the medals must have been listed separately. There is very little in the BM's database to represent this vast area of Sloane's collection and any analysis of the data held on the collections database will in turn be skewed, though researchers are not necessarily alerted to this when they search the BM's database. This reflect how data within systems, both historical and contemporary, has been input from different human sources with varying degrees of bias and accuracy. If projects do not acknowledge, or are not aware of this, analyses, and indeed findings, may be flawed, skewed and biased as a result (Institute of Museum and Library Services 2018; Jones 2022, 7).

2.2 Collection Data Atlas

The term 'Data Atlas' is predominantly used within the domains of computer science and information science and can take several forms usually influenced by domain or project characteristics. Consequently, there is no single, universally agreed definition of the term 'Data Atlas'. It has been used across several digital projects as a high-level abstraction to enable exploration, analysis, and synthesises, or visualisation, of usually disparate datasets.[3] However, it has also been used within a single domain of homogeneous datasets, as in the case of the *Atlas of Digitised Newspapers and Metadata*. This atlas, part of the *Oceanic Exchanges* project examined in depth the metadata of digitised newspapers across 10 international datasets and was created in response to the need for standardisation to enable meaningful connection across different datasets (Beals et al. 2020). The atlas provided a comprehensive visualisation of all the metadata fields and established a set of mappings that showed how metadata fields related to one another across different datasets.

Wong 2102 defined 'Data Atlas' as a compendium of diverse data objects, including maps, lists tables, illustrations, or analysis, resulting in a total overview of an organisation's information systems. The study proposed the development of an atlas and described its development methodology as comprising three key stages: the organisation of a multidisciplinary team; identification of information in all repositories and systems; and use of findings to build the atlas (Wong 2012). Similarly, the "Virtual Laboratory of neuroscientific data" presented the design and development of a technologically competent but conventional system architecture, entailing the term *Data Atlas* as an abstraction of the collective relation of the two main layers of the system architecture (Munir et al. 2015).

The role of a *Data Atlas* has also been understood as an interactive web platform for standardising data collection from Public Libraries. This allows for the organisation of measurable results under the Common Impact Measurement System (CIMS), with the aim of enabling public libraries to quantify their impact on individuals and communities beyond the metrics of services. The example differs from those above, acting more like a framework of data collection across seven areas of the CIMS: inclusion, culture, education, communication, economic development, health, and governance (Schrag 2015). Similarly, within the domain of biodiversity conservation, *Data Atlas* has been situated as a standard method in biodiversity fieldwork, holding strong ties to the domain of biological observation, and

3 See Beals et al. 2020; Wong 2012; Vroom 2019; Schrag 2015; Robertson, Cumming, Erasmus 2010; Parimbelli et al. 2022; Munir et al. 2015

providing a well-defined procedure of data collection at the point of field species observation (Robertson, Cumming, Erasmus 2010).

The term *Data Atlas* also appears in archaeological fieldwork data, as a tool to archive, store, link, and make accessible data which combines artefacts, written sources and pictorial evidence as information sources (Vroom 2019). It is also applied to an information system produced by PERISCOPE partners of integrated data about the COVID-19 Pandemic and its effect on health, economics, policy-making and society (Parimbelli et al. 2022). This atlas makes data readily available to the research community, decision makers and the public as a means of amplifying research and its impact, acting as a central solution for exploration and analysis of georeferenced data.

Table 1 The five characteristics of data atlas development as they appear in literature across 7 reviewed projects

	Multidisciplinary Team	Inventory quality	Review & Inspection	Data Collection	Integration & Organisation
Beals	×	×	×		×
Munir			×	×	
Parimbelli	×		×	×	
Robertson		×	×	×	
Schrag	×	×	×	×	
Vroom		×	×		×
Wong	×	×	×		×

Across all definitions of *Data Atlas* discussed above, we can apply a fine distinction between approaches focused on a data collection framework that drives implementation (Robertson, Cumming, Erasmus 2010; Parimbelli et al. 2022; Munir et al. 2015), and those that enable integration, mapping or organisation of pre-existing data or resources under a commonly understood arrangement.[4] These two distinct approaches are summarised in Table 1, where the first column reflects data collection, and the second represents integration and organisation. The table further summarises the projects in terms of key characteristics, including whether they engage a multidisciplinary team, use the Data Atlas as an inventory instrument, or employ it for review and inspection of data landscapes.

In recent years, there has been increasing interest in the creation of methodological tools that can increase transparency around the creation and use of datasets within a range of digital projects, from AI to Natural Language Processing (NLP) and Machine Learning

4 See Beals et al. 2020; Schrag 2015; Vroom 2019; Wong 2012.

(ML) (Mitchell et al. 2019). This has led to the adoption of systematic documentation approaches aimed at transparency and accountability by gathering information about dataset motivation, composition, collecting, pre-processing, labelling, intended uses, distribution and maintenance (Bender, Friedman 2018). By operating at the level of datasets such documentation approaches increase understanding of the contents and uses of datasets for internal and external project stakeholders, aiming to mitigate potential bias, overgeneralisation, or exclusion within project results. Whilst sharing the motivation to increase understanding of datasets used within digital projects, our approach to documentation of cross-institutional collection-as-data environments differs in a key respect. Rather than working at the level of the individual dataset alone, our approach focuses also on entities (institutions or digitisation projects) holding physical and digital data. This aims to augment our understanding of complex, cross-institutional data environments and to manage complexity within highly heterogeneous and immense datasets, resulting in better decision-making, particularly when navigating large and dispersed collection landscapes. More broadly, this approach as discussed in the section below offers a methodological instrument that other collections-as-data projects can adopt as an intentional data audit stage, informing key decisions about data aggregation, prioritisation, and access.

3 Methodological Framework

We propose *Collection Data Atlas* as a documentation paradigm and methodological instrument for the panoptic representation of cultural heritage collections, which aims to support understanding and decision making with regards to the status, scope and accessibility of collection and sub-collection datasets. Within the context of large-scale aggregation projects, our definition holds specific relevance for data audit and prioritisation, the decision-making process for selecting data to be included (or excluded) from aggregation, digital tools, techniques and workflows created for projects. Concurrently, it is a trans-institutional information tool, integral to efforts to comprehend the history of dispersed collections. It provides an understanding of data relating to a collection beyond the parameters of institutional holdings, in turn giving a deeper understanding of the documentation that has shaped knowledge of a cultural heritage collection over its history.

Drawing on the state-of-the-art review discussed above, we realise the following key characteristics that crosscut and give rise to the otherwise many and varied instances of Data Atlases presented above **[tab. 1]**: *Multidisciplinary Team*, *Data Atlas Metaphor, Review*

and Inspection, and *Inventory Quality*. The *Multidisciplinary Team* brings together experts, human intervention, consultation, and the practice of iterative development. The *Atlas Metaphor* refers to the definitions of the units used within the framework and can either hold a Geo-referencing connotation, an all-encompassing view, or a high-level unification. The *Review and Inspection* characteristic of the atlas facilitates decision-making and inspection of a collections landscape to aid understanding about the origin, availability, use and other relevant domain characteristics of interest. The *Inventory Quality* captures the order of arrangement of the data and resources and how they are grouped and labelled, as well as how complete and consistent is the inventory. The following sections address the four key elements of the methodology based our experience developing a Collection Data Atlas for the Sloane Lab project.

3.1 Multidisciplinary Team

The pivotal role and significance of a multidisciplinary team in the development of a Collection Data Atlas is recognised as integral to its successful creation. Multidisciplinary and cross-institutional teams can facilitate the creation of comprehensive instruments that foster shared understanding of a complex historic collections' landscape among team members and disciplines.

Crucial to this is the collaborative process, supporting a shared understanding of classification, use of language, definitions, membership, and the overall scope of the atlas.

Our initial conception of an atlas arose from the need to sketch a list of relevant data resources for ingestion and unification within the data aggregation environment of the project (i.e. the Sloane Lab Knowledge Base). Previous studies focusing on the original Sloane collection and historical manuscripts served as the starting point for generating a list of valuable resources (MacGregor 1994; Nickson 1988; Walker 2022). The first iteration of the atlas collated information about datasets and helped the multidisciplinary team to realise their breadth, availability and state. Further iterations led the team to the recognition of a broader landscape of resources beyond the historical set of the original Sloane Manuscript Catalogues, including collection guides, body of objects and other auxiliary and tertiary resources. The involvement of a multidisciplinary team in the design process of the atlas promoted shared understanding between the Technical Team (consisting of computer, data and information scientists) and the Collections Team (consisting of historians and museum curators), making both individual members and their institutions aware of the broader landscape of the Sloane collection.

3.2 Atlas Metaphor

The term 'atlas' often reflects the idea of all-encompassing and comprehensive representation of a world. In its conventional sense, an atlas conveys spatial and geographical information, carrying maps, charts, graphical representations and associated information like place names, statistics, and other textual descriptions. Geographically referenced datasets containing demographic and environmental information are commonly referred to as a data atlas (Robertson, Cumming, Erasmus 2010). However, the term data atlas can also have a metaphorical meaning in the context of information and data science, signifying a comprehensive and holistic representation of a system or a data framework. In this sense, data atlases draw upon the traditional concept to describe modules, components, and attributes of synthesised representations, offering a rich and multidimensional view of complex systems and data structures.

The atlas metaphor as a comprehensive representation of a 'world' is a highly versatile and abstract concept, making it particularly well-suited for integration within the diverse field of heritage and critical heritage studies. Central to this metaphor is the notion of continents as main sections that hold material and curate the vast scale of a collection. Its inherent abstraction enables flexible adaptation to specific requirements of diverse projects, ensuring an efficient approach to fit the complexities of a cultural heritage collection landscape.

3.3 Review and Inspection

A Collection Data Atlas fundamentally aims to help users understand and navigate complex collection landscapes using visualisations that support decision-making regarding data modelling, aggregation, ingestion, and resource prioritisation. Accordingly, the atlas should make affordances that balance richness of information and simplicity of representation. The high-level organisation of a Data Atlas should allow the grouping of collections under major categories like institutions and projects, with further aspects of collection particulars communicated through arrays of additional information. The combination of collection grouping and array casting of attributes enables a robust framework for insightful review and inspection. It supports users to select and scrutinise collections based on their distinctive characteristics, facilitating a comprehensive understanding of their origin, availability, and utilisation. Furthermore, this can accommodate evolving changes to meet the growth of collections, modifications in their characteristics,

and any further digitisation of the collection and its related documentation.

3.4 Inventory Quality

A Collection Data Atlas should support comprehension and navigation of complex and extended data landscapes. Thus, an inventory quality that can communicate organisation and membership across various collections and units is vital not only for the use of an atlas but also for its maintenance and growth. During the development of the Sloane Lab Data Atlas, we realised five design principles to ensure the inventory quality of the atlas namely *Clarity*, *Consistency*, *Analytical Arrangement*, and *Flexibility*.

Clarity seeks to prevent user confusion through cognitive overload and to support analysis and decision making based on a commonly shared and understood set of semantics about the type, state and availability of collection dataset. This is further reinforced by the Data Atlas Taxonomy discussed in the section below that supports *Consistency* via a hierarchical organisation and grouping of the various levels of item granularity. *Analytical Arrangement* refers to the use of an array of features per type of collection to ease the inspection and examination of individual units, allowing users to gain insights at a more granular level. Such arrangements reflect format, level of digitisation, digital availability, physical location and other useful attributes of items. Maintaining *Flexibility* allow accommodation of nuanced adjustments critical to the all-encompassing aims of an atlas, ensuring that it remains responsive to evolving needs and diverse requirements. This is achieved through an adjustable array of features and attributes that can be tailored to each item type, whether physical or digital. However, while catering for flexibility, there is a potential trade-off with design economy as it may lead to item replication.

Our approach focuses on the acquisition and collective view of datasets within the Sloane collection with the aim to integrate datasets into the Sloane Lab's knowledge base. In this process we define the scope and accessibility of these resources and identify their digitisation status and format, which enable us to design suitable data mapping and ingestion methods. Furthermore, our objective is to provide a methodological abstraction, particularly valuable to the field of DH, to support future projects in managing and comprehending complex landscapes of collections dispersed across, various institutions and systems of varying accessibility status.

4 The Sloane Lab Data Atlas

Building on the Collection Data Atlas method, we implement the Sloane Lab Data Atlas to provide a comprehensive view of the historical catalogues and modern datasets related to the Sloane collection, which are currently dispersed across various national institutions and individual projects. The implementation delivers a series of design affordances that balance the richness of information with ease of use across a complex cultural heritage collection landscape. It is also scalable, allowing continuous improvements and expansion as our understanding of the boundaries of the Sloane Collection develops. The Sloane Lab Data Atlas is realised as a tabular visualisation in the form of a Microsoft Excel Spreadsheet, because this allows for continuous improvement and expansion, whilst offering a range of options for formatting and high-resolution exports. Critically, Excel files are common tools that scholars and practitioners are familiar with (e.g. for searching, filtering, exporting data), flattening the learning curve for end users. The atlas **[Appendix]** is also scalable, allowing continuous improvements and expansion as our understanding of the boundaries of the Sloane Collection develops and has been made available on GitHub[5] as an open-source project and also directly available in PDF format[6] from the Sloane Lab portal.

Our approach focuses on integration, with the overarching aim of bringing together the many records, historical and modern, that describe the Sloane collection within a unified information environment called the Sloane Lab knowledge base. To achieve this, we establish the scope and accessibility of these records and identify their digitisation status and format in order to design suitable data mapping and ingestion methods. Furthermore, our objective is to provide real word case and working example to support future projects in managing and comprehending complex landscapes of collections dispersed across various institutions and systems with varying levels of accessibility.

4.1 Data Atlas Taxonomy

The Sloane Lab Data Atlas represents a diverse range of resource types and items, each having distinct forms, sizes, purposes, and granularity levels. Such resources expand beyond the original

5 See Github project page at `https://github.com/sloanelab-org/data-atlas`.

6 See PDF version of the of the Sloane Lab Data Atlas at `https://sloanelab.org/wp-content/uploads/2024/09/Data-Atlas-Vector.pdf`.

e-ISSN 2724-3923

Sloane manuscript catalogues, including historical resources such as collection guides, correspondence, minutes from when Sloane served as Secretary of the Royal Society (1693 to 1713), as well as modern resources such as digital surrogates, digitally born documents and databases. The Data Atlas taxonomy is designed for organising and categorising these resources based on their inherent characteristics and relationships. It is a hierarchical, structured framework designed to facilitate consistent grouping and provide clear definitions, enabling a comprehensive understanding of the breadth and depth of items, aiding classification, and supporting scalability. Consequently, the taxonomy serves as a system that allows the development of a shared language and structure to support communication and analysis of the participating resources within the Data Atlas.

The principal and foundational concept within the taxonomy is the 'Collection Unit' which refers to a physical or digital entity in a collection. It is not an abstract notion but rather a distinct entity characterised by specific attributes that are instantiated for both physical and digital items, including but not limited to size, location, degree of digitisation, transcription type, availability, accessibility, metadata schema, and programmable access. The notion of the collection unit originates from the NHM's *Join the dots* collections assessment exercise, where collections are arranged into discrete units that reflect how curators organise, index and work with their collections (Miller 2020). Our definition is elastic to allow the use of the Collection Unit as a building block to create a visual representation of the historical and contemporary Sloane collections. The taxonomical arrangement of the Collection Unit is presented in Figure 1.

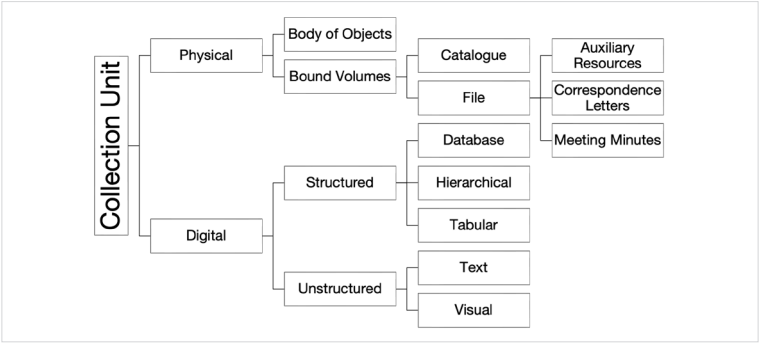


Figure 1 The Data Atlas Taxonomy

The Collection Unit may be Physical or Digital. Physical units are tangible and characterised by materiality. They can be touched or observed. Because the Sloane collection is historical, most units

contained in the Data Atlas are physical, for example, manuscript catalogues, collections of printed books, artifacts, artworks, specimens, etc. The digital surrogates of physical units, generated through a digitisation process such as scanning, photography and OCR (Optical Character Recognition) are considered attributes intrinsic to the physical unit rather than independent digital entities. This taxonomic design maintains a crucial distinction: between digital records of physical items (surrogates) and digital records about physical items (eg. representations and digital born material). This separation ensures that surrogate and born-digital materials remain distinct categories. The rationale reflects the atlas's primary focus on organising datasets and records of collections (i.e. the data records themselves) rather than the physical objects they describe.

Digital Units are retro-digitised or born-digital entities, either separate and distinct from any physical unit in the collection or born out of a digitation process, containing distinct and enriched attributes that augment and add value to the physical unit beyond its original form. An example is the fully searchable transcriptions in Text Encoding Initiative (TEI) format of the Sloane manuscript catalogues produced by the Enlightenment Architectures project (Enlightenment Architectures 2020). They introduce a set of important features and attributes such as Data Schema, Interactive Edition, Programmable Access, and Digital Location that are distinct from those associated with the original Physical Unit (i.e. manuscript catalogue). Similarly, the Sloane Letters Project[7] as a distinct digital collection unit, extends the volumes of Sloane correspondence held at the BL by adding summaries of letters, notes, item information and descriptive metadata (Sloane Letters n.d.). Inevitably, the differentiation between Physical and Digital Units may result in duplications throughout the Data Atlas, especially for collection units involved in extensive digitisation projects that yield substantial and enriched outputs. To establish a connection between these separate units originating from the same historical resource, the attribute *Origin* is introduced for digital units. This ensures a link between the distinct manifestations of the unit across different continents of the data atlas.

Additional categories are incorporated into the taxonomy to further classify Physical and Digital units. The Physical Unit is divided into Body of Objects and Bound Volumes. The Body of Objects consists of individual tangible items such as artefacts, artworks, specimens, etc. which are collectively understood as a curatorial unit but have not been arranged into a written catalogued form. Conversely, Bound Volumes are tangible objects, often comprised of written material bound together into single or multiple volumes or books, including

7 See project website The Sloane Letters Project. `https://sloaneletters.com/`.

materials that adopt the form of volumes or books, such as herbaria, where the written component supplements plant specimens pressed into the folios of a volume.

Bound Volumes are further categorised into Catalogue and File. The Catalogue follows the conventional definition of a systematic organisation of items arranged in a specific order, typically accompanied by descriptive information, providing details about the item's attributes, specifications, or characteristics. Conversely the File, includes all written components usually produced for record-keeping purposes that have a structured collection of information but are not catalogues. Files can be further classified into Meeting Minutes, Correspondence Letters, and other Auxiliary Resources.

The Digital Collection Unit is classified into Structure and Unstructured Units. The Structured Unit is organised and formatted according to a predefined schema, which may be well-defined and standardised. Structured units typically establish relationships and properties of data elements, making them amenable to querying and analysis using computational methods. Further categorisation of the structured units includes all types of Databases (i.e. relational and NoSQL), Tabular datasets such as spreadsheets, and Hierarchical XML data structures such as TEI files.

The Unstructured Unit does not offer the same level of structure and organisation to data as the Structured Unit, but it can be complex and rich in information. Such units are digitally born and can be further divided into Document and Visual. The Text is predominantly a text file that may contain figures and other pieces of information arranged in tabular format, available as a Word document, PDF file, Blog post and similar. The Visual unit is presented in form of images, maps, 3D models and other visualisations. Such visualisations may be the result of a tertiary implementation aimed at digitally reconstructing and preserving physical resources.

4.2 Data Atlas Continents

The Sloane Lab Data Atlas is arranged in rectangular panels, which are referred to as continents as per the atlas metaphor. Each continent visualises an institution or a major project that holds Collection Units related to the Sloane Collection. Collection units are visualised as different rows, which are grouped using appropriate categories based on the Data Atlas Taxonomy. To visualise different manifestations of the same Collection Unit (e.g. digital surrogates of a historic manuscript catalogue), a combination of colour coding and the use of separate columns within the same row is used. Footnotes clarify duplication of units appearing in more than a single continent. For example, the TEI manifestations of all collection units visualised

on the continent *Enlightenment Architectures* are also visualised on the Natural History Museum (NHM), the British Museum (BM) and the British Library (BL) continents respectively.

The section below outlines the details of the Sloane Lab Data Atlas, which organises both physical and digital collection units under distinct atlas 'continents'. Four institutional continents and three project continents are discussed whilst the details of an eighth continent, *Sloane Manuscript Catalogues Today* are presented, which aggregates all manuscript catalogues located across various heritage institutions.

4.2.1 Institutional Continents

The Institutional continents of the Sloane Lab Data Atlas include collection units from the NHM, BM, BL, and the Royal Society, with all institutions listing both Bound Volumes and Groups of Objects, except for the Royal Society, which lists only Bound Volumes. The NHM continent **[fig. 2]** holds 20 original Sloane manuscript catalogues, mostly related to fossils, insects, minerals, vegetables, and other specimens. It also contains 6 collection guides, including 3 volumes of the John Ray Historia Plantarum, 265 Sloane Herbarium Volumes, and 2 additional 'Bound Volumes' attributed to Leonard Plukenet (1641-1706) and James Petiver (1665-1718). The material has varying levels of digitization, transcription, and availability, as shown on the map. In addition, the NHM continent contains various contemporary datasets, including algae, fungi and plants, insects, invertebrates, palaeontology, mineralogy, and vertebrates.

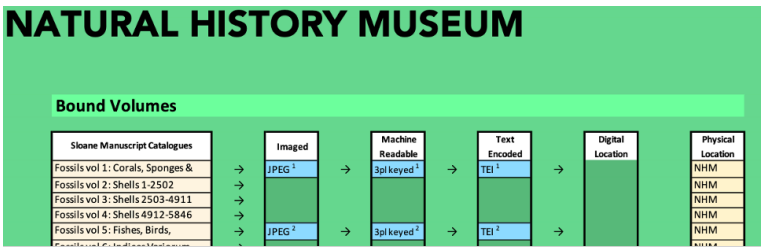


Figure 2 Part of the Natural History Museum dataset holding the original Sloane manuscript catalogues

The BM continent **[fig. 3]** includes four volumes of original Sloane manuscript catalogues, primarily focused on physical objects such as gems, cameos, amulets, and various other miscellaneous items. These manuscripts have varying levels of digitization and transcription and are physically housed at the British Museum. In addition, the continent organises approximately 15,000 objects from the BM's

Sloane collection across the eight curatorial departments where they are currently held: Prints & Drawings, Britain, Europe and Prehistory, Greece and Rome, Middle East, Asia, Coins and Medals, Africa, Oceania and the Americas, and Egypt and Sudan.

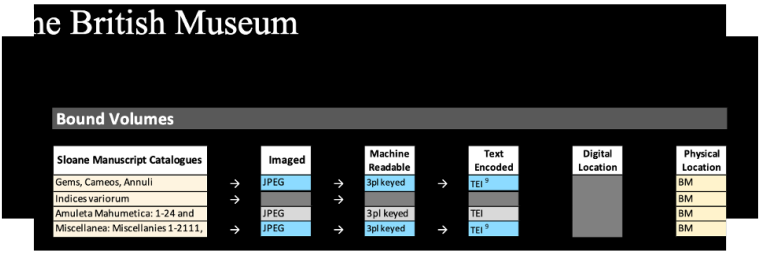


Figure 3 The British Museum dataset holding the original Sloane manuscript catalogues

The BL institutional continent **[fig. 4]** contains 21 volumes of original Sloane manuscript catalogues, each with varying levels of digitisation and transcription, detailing the large collection of printed books and manuscripts gathered by Sir Hans Sloane. At the time of his death, Sloane's library was estimated to hold 50,000 volumes, of which approximately 47,200 were printed items. His manuscript collection, comprising around 5,200 items, covers a wide range of subjects, including medicine, alchemy, chemistry, botany and horticulture, exploration and travel, mathematics, natural history, magic, and religion. The entirety of Sloane's collected printed books and manuscripts is catalogued in the British Museum's online catalogue and organized into relevant datasets. Additionally, the BL continent includes two separate entries for datasets containing approximately 1,400 of Sloane's correspondence letters.

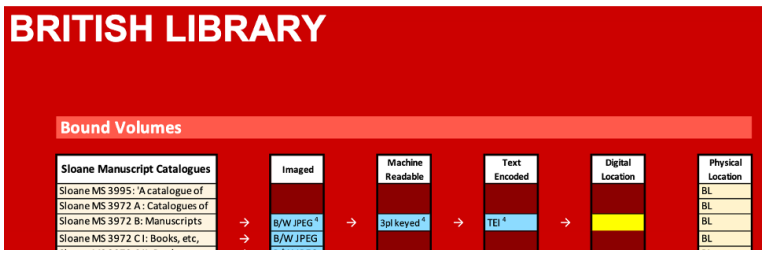


Figure 4 Part of the British Library dataset holding the original Sloane manuscript catalogues

The Royal Society's collection **[fig. 5]** includes five separate units of meeting minutes spanning the years 1686 to 1711, containing summaries of Royal Society meetings, documenting discussions on

experiments, publications, and natural curiosities. The majority of the manuscripts were compiled by, or on behalf of, Hans Sloane, who served as Secretary of the Royal Society from 1693 to 1713, and later as President (1727-41). The minutes from 1686 to 1711 have been digitised, imaged, and transcribed, and are available online through the Royal Society's.

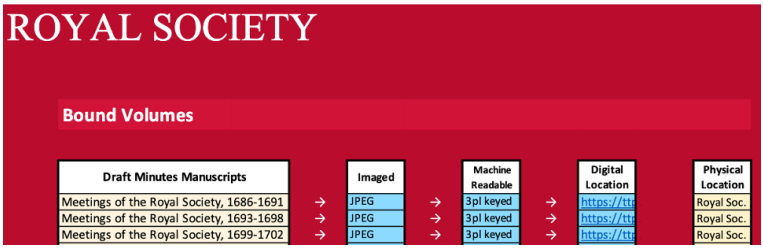


Figure 5 Part of the Royal Society dataset holding the original Sloane manuscript catalogues

4.2.2 Project Continents

The project continents contain collection units related to the work of significant research projects that investigated, organised, and digitised Sir Hans Sloane's original manuscript catalogues and other relevant materials from his collection. The Adam Matthew Library's digital data includes the Commonplace book manuscripts collected by Sloane, all of which are imaged and transcribed into machine-readable formats. The Sloane Letters Project is a database encompassing thirty-eight volumes of Sloane's correspondence held at the British Library (MSS 4036-4069, 4075-4078). It includes descriptions and metadata for Sloane MSS 4036-4035 and 4075, along with some letters and transcription from MSS 4054-4055, 4066, 4068-4069, and 4076. The Enlightenment Architectures continent includes the Text Encoding Initiative (TEI) outputs of the project, funded by the Leverhulme Trust. The project was a collaboration between the British Museum and University College London, with contributions from the British Library and the Natural History Museum. The project transcribed and created open-access, TEI-encoded remediations of five of Sloane's catalogues: two volumes on 'fossils', one volume of printed books and ephemera, one volume of 'miscellanies', and one cataloguing his collection of manuscripts (Ortolja-Baird et al. 2019, 44).

4.2.3 The Sloane's Manuscript Catalogues Today

The 'Continent' aggregates and indexes the surviving Sloane manuscript catalogues, which have been used and annotated by subsequent curators for three centuries now. During his lifetime, Sir Hans Sloane produced catalogues of his entire collection, which by 1725 had grown to 32 in number. The size of the catalogues varied, with some catalogues listing thousands of items and others being relatively short and probably combined or even bound by Sloane into single larger volumes. At his death in 1753, the numbers of items in each of his 32 catalogues had increased substantially, as recorded by the list transmitted to his executors after his death (MacGregor 1994, 28-9). The original catalogues continued to be used over the next centuries by museum curators. As curatorial departments grew, merged and changed, catalogues were copied, split up and recombined for ease of reference. In 1998 Peter M. Jones published the book *A preliminary checklist of Sir Hans Sloane's catalogues* which set the basis of a fully descriptive listing of Sloane's Catalogues (Jones 1988), and together with MacGregor A. in 1994 the book *Sir Hans Sloane, Collector, Scientist, Antiquary, Founding Father of the British Museum*. Both list 31 catalogues of the Sloane collection in total. The Jones/MacGregor numbering of catalogues (i.e. 1-31) reflects the complex bibliographic and curatorial history of the collection, which in some cases led to the creation of new catalogues for ease of reference, based on the organizational structure of heritage institutions. In other cases, multi-volume catalogues, such as the catalogue of printed books and manuscripts (12 volumes) were assigned a single number, while other catalogues, originally bound together by Sloane, such as Minerals (5 volumes) and Fossils (6 volumes) were split into individual numbers. Jones and MacGregor also included in their numbering volumes that were indices to various hand-written catalogues created by Sloane's assistant Thomas Stack (d.1756).

The Sloane Lab Data Atlas diverges from the Jones/MacGregor numbering system by listing Sloane's original catalogues by individual volume when these are predominantly by him as a catalogue of his collection **[fig. 6]**. However, the Jones/MacGregor numbering is maintained as a concordance to facilitate reference to earlier projects and publications. This explains why the number of catalogues in the atlas exceeds the 31 catalogues listed by Jones and MacGregor, while also omitting some catalogues they included.

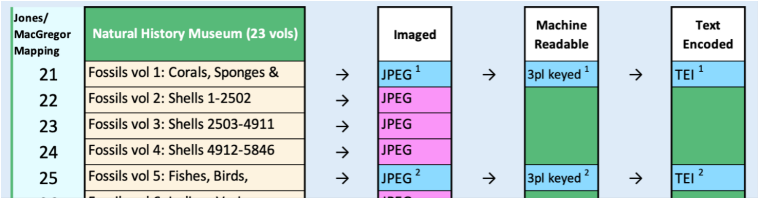


Figure 6 Part of the Sloane's Manuscript Catalogues Today held by the Natural History Museum (NHM), illustrating the MacGregor mappings

4.3 Data Atlas Attributes and Organisation

The atlas implements the inventory design principle by organising collection units by continent and arranging the continents into distinct sections while using colour coding, grouping, and attribute assignment to further enhance structure and clarity. Each continent is assigned a unique background colour to distinguish between institutions or digitisation projects and is organised into sections based on collection unit types, such as 'Bound Volumes' and 'Group of Objects' **[Appendix]**. Sections contain one or more tables, representing the different groups of collection units, with each table row containing an individual collection unit. The table columns represent various attributes assigned to collection units: for Bound Volumes these distinguish between a physical manuscript catalogue and its digital surrogates. For Group of Objects they reflect attributes related to state, size and availability. Moreover, cell colour is used to represent further attributes of collection units related to format and current availability, such as digitally available (blue), digital in the process of becoming available (grey), tangible objects (ivory), and online URL (yellow). Figure 7 illustrates an example arrangement of British Museum collection units, divided into 'Bound Volumes' and 'Group of Objects'. The different levels of digitisation of historical manuscripts are shown as Imaged JPEG, Machine Readable triple keyed transcription, and Text Encoded in TEI while the contemporary, databased catalogue, of Prints and Drawing contains a separate set of attributes about the state, size and availability of the collection.

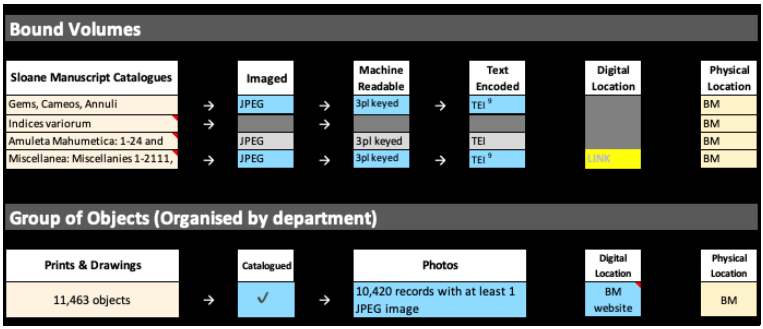


Figure 7 Part of the British Museum atlas continent illustrating organisation of collection units for Bound Volume and Groups of objects

5 Discussion and Reflections

5.1 Benefits and Limitations

The Sloane Lab Data Atlas presents a panoptic visualisation of a complex bibliographic and data environment and represents the first time that the contemporary data and institutional landscape of Sloane's collections has been brought together. As a 'collections as data' project, clarity around the scope, volume, availability, nature and digital accessibility of the project's data is of central importance to the Sloane Lab, its team, partners and stakeholders. The Atlas delivers this, bringing together the expert knowledge of curators and information scientists to act as an organisational tool for infrastructure development. It has aided the design of appropriate data mapping, modelling and ingestion approaches for the Sloane Lab project and wider development of its Knowledge Base. It has enabled partner institutions to identify data absences in their digitised collections that led to revisiting digitisation pipelines to cover such absences. Furthermore, it has given cultural heritage institutions and universities alike an important further case study of the importance of institutional collaborations between curators, academics and technical teams.

Most importantly, the Collection Data Atlas is a methodological instrument that can be beneficial to large-scale digital projects that deal with complex data environments in three distinct ways. Firstly, it augments a project team's understanding of different information systems, what data they hold, and their level of digitisation and availability is. Secondly, it facilitates the scoping out of complex data environments and improves project-wide decision-making, namely the prioritisation of datasets for ingestion into, for example, a Knowledge

Base and the effective budget allocation concerning digitisation for a given project and institutional partners alike. And thirdly, it allows for these complex data environments to be better communicated with stakeholders such as funders, institutional collaborators and with end-users.

These three key benefits are embodied by the Sloane Lab Data Atlas. However, to achieve and maintain this balance, a series of accommodations were made. Whereas the high-level organisation of the Data Atlas is well defined, at the lower level (i.e., closer to the data) the conventions of the Sloane Lab Data Atlas are purposely flexible. A prime example of this is how the different columns are understood differently depending on the section, as explained above. Moreover, whilst the Sloane Collections are vast, the Sloane Lab Data Atlas visualisation is agile and scalable to accommodate future changes and modifications. Accordingly, though a graphic visualisation made using a design tool may be subjectively more appealing, the Sloane Lab Data Atlas has been developed using Microsoft Excel, balancing these desiderata. Added to this, with the current approach we have encountered the issue of double instancing. For the purposes of simplicity, the Sloane Lab Data Atlas addresses this issue with explanatory footnotes, something that a technically strict visualisation may not allow for.

A significant future avenue for the development of the Data Atlas would be to bring into question the unnamed, enslaved people who related to some of the objects in Sloane's collection (see Introduction) and thus with the historical records of it that are mapped by the Atlas. In their 2022 paper on the question of silence and bias in the early-modern archive, Ortolja-Baird and Nyhan extracted the names of c. 3,000 people and 600 geographical locations from just two of Sloane's Manuscript Catalogues, noting how this quantitative examination of historical records ":attest a discrepancy of people and places in colonial and imperial contexts, and hints at just how many persons are absent from Sloane's network" (Ortolja-Baird, Nyhan 2022, 855). As the authors note, the individuals most named within Sloane's catalogues are British or European individuals. This raises a question to be addressed in future work of how the Data Atlas can make visible the quantity of marginalized voices who were central to the way in which Sloane's knowledge of his collections was formed and catalogued within his lifetime, in turn influencing the cataloguing and curation of his collections within the institutions in which they are now held, and ultimately the digitisation of historical records. With each stage in this knowledge making process the polyvocal sources of Sloane's knowledge have been further silenced. Following Lawther, this question has the potential to bring people back to the centre of Sloane's collections, in turn addressing power dynamics and hierarchies that have been present within the

collection and its record for c. 340 years (Lawther 2023). This it only acts to emphasise the importance of recognising the genealogy of museum documentation (Goskar 2024), and the impact that the history of a collection or institution's documentation has on the future ability to identify provenance (MacDonald 2023, 317). The Collections Data Atlas addresses these challenges by providing a comprehensive documentation framework that clarifies the scope, volume, availability, nature, and digital accessibility of collection and sub-collection datasets, supporting this way both historical understanding and future provenance research.

6 Conclusion

In this paper, the Collection Data Atlas has been positioned as an instrument that enables data-driven researchers to work with historians, curators, bibliographic specialists and others to establish what is extant, how it has been recorded, and, crucially, who has created it. The atlas presented is thus an instrument for data collection, inventory and multidisciplinary exploration. In the applied instance (i.e. Sloane Lab Data Atlas) explored in this paper, it is also the first synthesis of its kind, and visual representation of the historical and contemporary collection records of Sloane, now dispersed across different institutions, information systems and infrastructures.

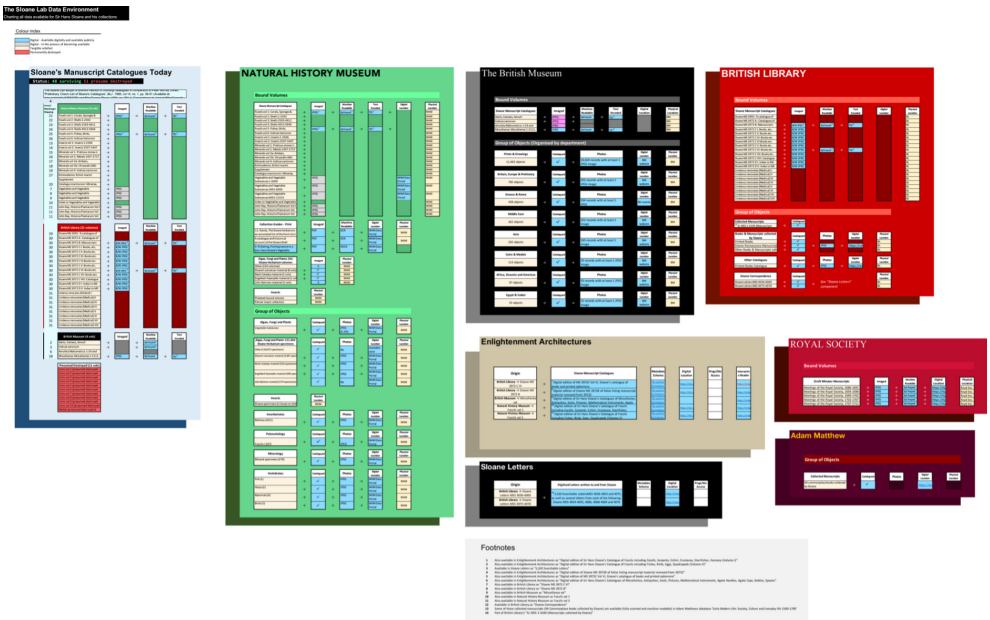
Underlying our discussion is an acknowledgement of three axes of the history of Sloane's documentation: firstly, the way in which Sloane and his contributors catalogued his collection within his lifetime; secondly, how the collection has been recorded and curated by the varying institutions it is now housed within and, thirdly, the digitisation practices by cultural heritage institutions and digital projects of the historical and institutional documentation relating to the Sloane collections. Sloane's collection was created through the economic, political and cultural processes of Britain's increasing global entanglements of the seventeenth and eighteenth centuries. As such, the question of who has created these historical records is central to further an understanding of the dynamics of power that have shaped knowledge around Sloane's collections, and the impact this has on the contemporary contexts within which his collections are located. This has been identified for a next step in the development of this tool. As it stands, the Sloane Lab Data Atlas is the first step in the wider project's development of new and better computational approaches to the detection and visualisation of these past entanglements ever present through gaps, biases and exclusion within the digitally available historical record. Likewise, following the transferability of the data atlas, we hope that it can be a tool that

can further conversations and computational interventions about this in other data-driven digital projects.

So too, the range of documents in which the Sloane collection is described and contextualised is vast. It extends beyond some 40 manuscript catalogues compiled by Sloane and his assistants and the records of the national institutions mentioned above. They include historical resources such as collection guides, correspondence, minutes from when Sloane served as Secretary of the Royal Society (1693 to 1713), as well as modern resources such as digital surrogates, digitally born documents and databases.

Appendix: Snapshot of the Data Atlas

The appendix presents the Sloane Lab Data Atlas in full deployment, comprising all eight data contents that contributed to the extended version of the Atlas for the data review purposes of the Sloane collection. The datasets include the Sloane manuscript catalogues as they appear today, the datasets held by the Natural History Museum, the British Museum and the British Library, as well as datasets that have been handled by individual projects including the Enlightenment Architectures, the Sloane Letters, the Minutes of the Royal Society and the Adam Matthew's digitisation project.



Bibliography

Beals, M.; Bell, E., Cordell, R.; Fyfe, P.; Russell, I.G.; Hauswedell, T.; Neudecker, C.; Nyhan, J.; Oiva, M.; Pado, S.; Pimentel, M.P. (2020). *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges*. Online: Figshare, Loughborough University. https://doi.org/10.6084/m9.figshare.11560059.

Bender, E.M.; Friedman, B. (2018). "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science". *Transactions of the Association for Computational Linguistics*, 6, 587-604.

Berger, M. (2023). "Between Canon and Coincidence". *Journal for Art Market Studies*, 1. https://doi.org/10.23690/jams.v7i1.147.

Caygill, M. (1994). "Sloane's Will and the Establishment of the British Museum". MacGregor, A. (ed.), *Sir Hans Sloane: Collector, Scientist, Antiquary, Founding Father of the British Museum*. London: British Museum Press.

Cornish, C.; Driver, F. (2020). "'Specimens Distributed': The Circulation of Objects from Kew's Museum of Economic Botany, 1847-1914". *Journal of the History of Collections*, 32(2), 327-40. https://doi.org/10.1093/jhc/fhz008.

de Boer, V. et al. (2012). "Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study". Simperi, E. et al. (eds), *The Semantic Web: Research and Applications*. Berlin: Springer, 123-34.

Dixon, C. (2023). *Sailing the Monsoon Winds in Miniature: Understanding Indian Ocean Boat Models*. London: British Museum Press.

Enlightenment Architectures (2020). "Enlightenment Architectures: Sir Hans Sloane's Catalogues of His Collections". *Reconstructing Sloane*. https://reconstructingsloane.org/enlightenmentarchitectures/2017/07/19/featured-content/.

Gosden, C.; Larson, F.; Petch, A. (2007). *Knowing Things: Exploring the collections at the Pitt Rivers Museum, 1884-1945*. New York: Oxford University Press.

Hauswedell, T.; Nyhan, J.; Beals, M.H.; Terras, M.; Bell, E. (2020). "Of Global Reach yet of Situated Contexts: An Examination of the Implicit and Explicit Selection Criteria That Shape Digital Archives of Historical Newspapers". *Archival Science*, 20(2), 139-65. https://doi.org/10.1007/s10502-020-09332-1.

Hodel, T. (2023). "Konsequenzen Der Handschriftenerkennung Und Des Maschinellen Lernens Für Die Geschichtswissenschaft. Anwendung, Einordnung Und Methodenkritik". *Historische Zeitschrift*, 316(1), 151-80. https://doi.org/10.1515/hzhz-2023-0006.

Hyvönen, E. (2012). "Publishing and Using Cultural Heritage Linked Data on the Semantic Web". *Synthesis Lectures on the Semantic Web: Theory and Technology*, 2(1), 1-159.

Institute of Museum and Library Services (2018). *The Santa Barbara Statement on Collections as Data (v2)*. https://collectionsasdata.github.io/statement/.

Jones, M. (2022). *Artefacts, Archives and Documentation in the Relational Museum*. Oxford: Routledge.

Jones, P.M. (1988). "A Preliminary Check-List of Sir Hans Sloane's Catalogues". *The British Library Journal*, 14(1), 38-51.

Lawther, K. (2023). *People-Centred Cataloguing*. http://www.kathleenlawther.co.uk/wp-content/uploads/2023/01/people-centred-cataloguing-final.pdf.

Leiden University (2023). *Archaeologist Martin Berger explores Latin American collections with an ERC grant*. https://www.universiteitleiden.nl/

en/news/2023/09/archaeologist-martin-berger-explores-latin-american-collections-with-an-erc-grant.

Luther, L. (2022). "Introduction: Digital Benin". *Digital Benin*. https://digitalbenin.org/documentation/introduction.

MacDonald, I. (2023). "Counting When, Who and How: Visualising the British Museum's History of Acquisition Through Collection Data, 1753-2019". *Journal of the History of Collections*, 35(2), 305-20.

MacGregor, A. (ed.) (1994). *Sir Hans Sloane: Collector, Scientist, Antiquary, Founding Father of the British Museum*. London: British Museum Press.

Miller, G. (2020). *Collection Unit Definition and Guidelines (From "Join the Dots" Collections Assessment Exercise)*. Natural History Museum. https://data.nhm.ac.uk/dataset/join-the-dots-collections-assessment-exercise/resource/c3c83129-27f7-4bb9-9d3f-e85f4d550f55.

Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T (2019). "Model Cards for Model Reporting" = *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. New York, NY: Association for Computing Machinery, 220-9. https://doi.org/10.1145/3287560.328759.

Modern Migrants Project Team (s.d.). "The Research Project". *Modern Migrants: Painting from Europe in US Museums*. https://www.modernmigrants.art/.

Munir, K.; Ahmad, KH.; McClatchey, R. (2015). "Development of a Large-Scale Neuroimages and Clinical Variables Data Atlas in the neuGRID4You (N4U) Project". *Journal of Biomedical Informatics*, 57, 242-62.

Nickson, M.AE. (1988). "Hans Sloane, Book Collector and Cataloguer, 1682-1698". *The British Library* Journal, 14, 52-89.

Oceanic Exchanges Project Team (2017). *Oceanic Exchanges: Tracing Global Information Networks in Historical Newspaper Repositories, 1840-1914*. https://doi.org/10.17605/OSF.IO/WA94S.

Nyhan, J.; Vlachidis, A.; Flinn, A.; Pearlman, N.; Carine, M.; Hill, J.; Humbel, M.; Jansari, S.; Sloan, K.; Pickering, V.; Valeonti, F. (2025). "Final Report - Sloane Lab: Looking Back to Build Future Shared Collections". *Towards a National Collection*. https://doi.org/10.5281/zenodo.14771754.

Ortolja-Baird, A.; Pickering, V.; Nyhan, J.; Sloan, K.; Fleming, M. (2019). "Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane's Early Modern Catalogues of His Collections". *Open Library of Humanities*, 5(1), 44.

Ortolja-Baird, A.; Nyhan, J. (2022). "Encoding the Haunting of an Object Catalogue: On the Potential of Digital Technologies to Perpetuate or Subvert the Silence and Bias of the Early-Modern Archive". *Digital Scholarship in the Humanities*, 37(3), 844-67.

Parimbelli, E.; Larizza, C.; Urosevic, V.; Pogliaghi, A.; Ottaviano, M.; Cheng, C.; Benoit, V.; Pala, D.; Casella, V.; Bellazzi, R.; Giudici, P. (2022). "The PERISCOPE Data Atlas: A Demonstration of Release v 1.2". Michalowski, M (ed) = *Proceedings of the 20th International Conference on AI in Medicine (AIME 2022)*. Cham: Springer International Publishing, 412-15.

Penn, M.G.; Cafferty, S.; Carine, M. (2018). "Mapping the History of Botanical Collectors: Spatial Patterns, Diversity, and Uniqueness Through Time". *Systematics and Biodiversity*, 16(1), 1-13.

National Gallery of Art. "National Gallery of Art Collaborates with Researchers to Analyse Permanent Collection Data". https://www.nga.gov/press/2019/datathon.html.

Padilla, T. (2017). "On a Collection as Data Imperative". *UC Santa Barbara*. https://escholarship.org/content/qt9881c8sv/qt9881c8sv.pdf.

Padilla, T.; Allen, L.; Frost, H.; Potvin, S.; Russey Roke, E.; Varner, S. (2019). "Always Already Computational, Collections as Data Final Report". https://doi.org/10.5281/zenodo.3152935.

Petch, A. (2002). "Today a Computerised Museum Catalogue: Tomorrow the World". Special issue, *Papers Originating from the MEG Conference 2001* of *Journal of Museum Ethnography*, 14, 94-9.

Petch, A. (2006). "Counting and Calculating: Some Reflections on Using Statistics to Examine the History and Shape of the Collections at the Pitt Rivers Museum". *Journal of Museum Ethnography*, 18, 149-56

Phillipson, T. (2019). "Collections Development in Hindsight: A Numerical Analysis of the Science and Technology Collections of National Museum Scotland Since 1855". *Science Museum Group Journal*, 12.

Punzalan, R.L. (2014). "Understanding Virtual Reunification". *Library Quarterly: Information, Community, Policy*, 84(3), 294-323.

Rother, L.; Mariani, F.; Koss, M. (2023). "Hidden Value: Provenance as a Source for Economic and Social History". *Economic History Yearbook*, 64(1), 111-42.

Robertson, M.P.; Cumming, G.S.; Erasmus, B.F.N. (2010). "Getting the Most Out of Atlas Data". *Diversity and Distributions*, 16, 363-75.

Schrag, T. (2015). "Building a Public Library Impact Data Hub: A Global Libraries 'Data Atlas' for Storytelling, Strategy Development, and Collaboration". *International Federation of Library Associations (IFLA) World Library and Information Congress (WLIC 2015)*.

Shoilee, S.B.A. (2022). "Knowledge Discovery for Provenance Research on Colonial Heritage Objects". *Doctoral Consortium at ISCW 2022 co-located with 21st International Semantic Web Conference*.

Siqueira, J.; Martins, D.L. (2022). "Workflow Models for Aggregating Cultural Heritage Data on the Web: A Systematic Literature Review". *Journal of Association for Information Science and Technology*, 73, 204-24.

Solà, M.C.; Korepanova, A.; Mukhina, K.; Schich, M. (2023). "Quantifying Collection Lag in European Modern and Contemporary Art Museums". *VINCI '23: Proceedings of the 16th International Symposium on Visual Information Communication and Interaction*, 39, 1-8.

Vroom, J.A.C. (2019). "Data Atlas of Byzantine and Ottoman Material Culture: Archiving Medieval and Post-Medieval Archaeological Fieldwork Data from the Eastern Mediterranean (600-2000 AD), Phase 1". *Research Data Journal for the Humanities and Social Sciences*, 1-12.

Wingfield, C. (2011). "Donors, Loaners, Dealers and Swappers: The Relationship between the English Collections at the Pitt Rivers Museum". Byrne, S.; Clarke, A.; Harrison, R.; Torrence, R. (eds), *Unpacking the Collection: Networks of Material and Social Agency in the Museum*. New York: Springer.

Walker, A. (2022). "Sir Hans Sloane's Books: Seventy Years of Research". *British Library Journal eBLJ*, Article 6.

Wong, W. (2012). "Mapping Your Way to Compliance with a Data Atlas". *Information Management*, 46(1).

Zaagsma, G. (2022). "Digital History and the Politics of Digitization". *Digital Scholarship in the Humanities*, 38(2), 830-51. https://doi.org/10.1093/llc/fqac050.