TOWARDS A NATIONAL COLLECTION

UKRI | **Arts and Humanities Research Council**

FIRST REPORT
DISCOVERY PROJECTS

Sloane Lab:
Looking back to build
future shared collections

Table of Contents

Authors

Julianne Nyhan (UCL), Hanna James (UCL), Andreas Vlachidis (UCL), Andrew Flinn (UCL), Nina Pearlman (UCL), Mark Carine (Natural History Museum), Jeremy Hill (The British Museum)

Executive Summary

How can galleries, libraries, archives and museums (GLAMs) use digital technology to link their collections together in ways that make it easier for different people, both specialists and interested publics, to find the information they want? How can digital technology help us tell new stories about what can be rediscovered and reimagined by linking collections? How can we make specialist users and members of the public more aware of the contested nature and histories of museum collections? What is the role of digital tools in foregrounding overlooked or hidden processes, like imperialism, colonialism, the slave trade, loss and destruction, that have shaped the national collection? Who gets to contribute to, and shape, research on how memory institutions can reach across their institutional boundaries, subject-specialities and even countries so as to better engage their varied audiences? And how can heritage institutions select the most useful technologies from the many that are available for digitising, releasing and interlinking their collections? These are the questions to which *the Sloane Lab: Looking back to build future shared collections* is responding.

Taking the collections assembled by Sir Hans Sloane as an exemplary case study that is expected to be transferable to the wider collections as data context, the project will unite surviving Sloane objects, records and knowledge across the Natural History Museum (NHM), British Museum (BM) and British Library (BL) from the original catalogues, also furthering understandings of the extent and provenance of Sloane's original 1753 collection, the founding national collection of the BM. Bringing Sloane's historical catalogues and present-day cataloguing systems together, linking to other digital collections, and learning how to better support different ways to search and accommodate any "curious or interested person", the project also addresses what is missing and contested. By working directly with local, regional, national and international heritage partners to devise trans-institutional methodologies for the integration of other collections with the Sloane Lab we hope to advance collections as data in the context of the holdings of UK GLAM. Our work will be crystallised in the Sloane Lab, where the reunified and extended digital collection will be made searchable and amenable to further analysis and interpretation using a range of digital tools and processes.

Our project, *Sloane Lab: Looking back to build future shared collections* has made excellent progress since its initiation on 1 October 2022. Turning firstly to summarise the essential project governance and set-up processes that have been completed: a full cohort of project staff has now been appointed, and all necessary Human Resources and management requirements and processes pertaining to their employment completed. UCL's Ethics Review Board has reviewed the project and granted its proposed plan of work the necessary approval; likewise, all data protection and data assurance processes have been set up and approved by the relevant UCL committee. The Collaboration agreement between UCL, the British Museum and Natural History Museum has been agreed upon in all parts aside from the licences to be attached to metadata, but we believe that we have now agreed on a mutually agreeable route with regard to this and that the collaboration agreement can be approved shortly.

This has been detailed and demanding work, which has taken a significant amount of project time and yet it is of the utmost importance because the agreement that is reached now will inform the path that Heritage Organisations like the British Museum will take with regard to the licencing of their metadata for the future digital national collection. Further collaboration agreements have been issued to the British Library and the one outstanding collaboration agreement, with metaphacts, is being drawn up

now by UCL's legal team. Next to this, a number of consultations with project partner organisations and with the Advisory Board have been pursued and are upcoming.

The research-work of the project is being pursued across three interconnected streams: the technical team, the participatory team and the collections team.

The technical team has now completed the setup and deployment of the Sloane Lab technical infrastructure, including the set up and deployment of servers, portals and schedules. Challenging work pertaining to the knowledge base development, the elaboration of a CIDOC-CRM data model and the alignment of the extant TEI-encoded Sloane structured data with this model has now completed its first iteration, allowing the data to be ingested into the knowledge base. Working to enhance the resolution of the knowledge-enrichment tasks that can be executed there, work on data enrichment, for example pertaining to people and places, is ongoing, along with Natural Language Processing and potentially crowd-sourced routines to automate aspects of this. So too, crucial tasks pertaining to front-end deployment and dissemination are well-advanced, including website Development, Data Atlas and Workflows and the mobilisation of additional capstone-data for the project, namely that of Sloane's annotated copy of John Ray's (1686–1704) *Historia Plantarum* that serves as the taxonomic index to his herbarium. Work now is on ensuring the effectiveness of this work but the necessity of abstracting and disseminating the technical processes and routines of the project is on our horizon.

The participatory team is moving forward in its tasks to engage diverse communities with the emerging Sloane Lab, to understand what expert and non-hegemonic communities wish to ask of a digital national collection and to understand the barriers and accelerants that currently exist for heritage organisations, of all sizes and thematic interests, to participate in the digital national collection. Following the completion of a wide-ranging literature review on the latter question, qualitative instruments have now been completed to elicit the state of the art from heritage and data professionals and institutions as to their envisaged capacity and readiness to participate in the foreseen digital national collection. Moreover, a detailed participatory plan for engaging in a wealth of knowledge co-creation activities with specialist and non-hegemonic communities has now been finalised, and a plan of implementation over the next two years is on course for agreement. Most distinctive about this plan is the extent to which it has tethered the participatory aims of the project to the technical elaboration of the Sloane Lab infrastructure and information modelling activities. Accordingly, it is expected that this work will result not only in the creation of new knowledge about what individuals and communities want from a digital national collection, but also that the methodologies and instruments being elaborated for this work can be disseminated and foment a 'participatory turn' in collections as data.

The collections team are now working closely with the technical team to carry out data review, catalogue data cleaning and enhancement, and to create a single data set of natural history specimens for data mobilisation. In the coming months, in addition to supervising community fellowships, the collections team will also lead co-authored publications on *Historia Plantarum* data mobilisation and cataloguing of the botanical collections (of course, co-authored papers will be written up by all of the three interconnected teams). This collections team will also contribute to the planned trans-institutional case study, which will not only create new knowledge about the Sloane collection but act as a test bed for the Sloane Lab. They will lead the research and preparation for the touring exhibition with a number of venues including Down County Museum. There will also be permanent display changes in the British Museum.

Abstract

The founding collection of the British Museum is a rich area to explore how we can reconnect dispersed heritage connections using state of the art technologies. This is because the British Museum's original 1753 founding collection of Sir Hans Sloane is now split across three different institutions (the British Museum (BM), Natural History Museum (NHM) and the British Library (BL)) and the digital information that describes this founding collection sits in the different institutions in a range of different systems that are not currently set up to talk to one another. By focusing on catalogue records, and the vast, remaining collections of Sir Hans Sloane, the Sloane Lab project is researching how we can work with interested communities and heritage organisations to link the present with the past so as to allow the currently broken links between Sloane's collections and catalogues to be re-established across the NHM, BL, BM (plus others that have relevant material).

The main outcome of our project will be a freely available, online digital lab (the Sloane Lab) that will offer researchers, curators and the interested public new opportunities to search, explore, and critically and creatively use and reuse digital cultural heritage.

Aims and Objectives

The aims and objectives that the Sloane Lab will deliver include:

- To devise automated and augmented ways that will be relevant to bringing together other collections in museums, galleries, libraries and archives, of mending the broken links between the past and present of the UK's founding collection in the catalogues of the British Museum, Natural History Museum and the British Library. For this, we will use the collection of Sir Hans Sloane (1660-1753) as a microcosm through which to explore the technical, infrastructural, conceptual, historical and social challenges faced in bringing together digital cultural heritage collections so as to help audiences use, learn and benefit from them.

- To deliver the public-facing, open access Sloane Lab of content, tools and applications, opening new ways for national digital collections to be researched, explored, interlinked and creatively engaged with.

- To facilitate richer, more critical understandings of the origins and development of museum collections by devising computational and conceptual approaches to detecting and exposing often-hidden processes like colonialism, empire and slavery that have shaped collections and their classifications. Sloane's collection was created through the economic, political and cultural processes of Britain's increasing global entanglements of the 17/18th century, to which the infrastructure for a 21st-century national collection must respond.

- To evaluate the role of AI, Digital Humanities, Data Science and historical collection records in the infrastructure of the future national collection, disseminating our findings globally as open source and reusable demonstrators and best practice documentation.

- To engage in ongoing, participatory research with diverse audiences, project partners, and the wider heritage sector to elaborate a digital cultural heritage participatory research paradigm and methodology that accommodates a plurality of perspectives and voices and does not limit the publics to interacting with the national collection as curators, librarians, archivists, computer scientists and others have modelled it in existing platforms.

- To work with Heritage Institutions to identify factors like resources or investments that constrain an organisation's ability to contribute to national collection infrastructures and content and deliver policy recommendations for future digital heritage capacity and infrastructure investment.

Partnership structure

Collaborators

The Sloane Lab is led by University College London (UCL) which oversees the research and technical trajectory and alignment (including ethical oversight and data protection) of the project as a whole, along with administrative and budgetary management and reporting on the project to AHRC and the TaNC Directorate. Further particulars of UCL's contributions to every work package of this project are given below. Project collaborators are the British Museum (BM) and Natural History Museum (NHM), where Co-Investigators Jeremy Hill (BM) and Mark Carine (NHM) oversee the execution of their respective institutions' deliverables for the project and supervise the Sloane Lab-funded project staff hired by their institutions. The NHM team work closely with the technical team to carry out data review, catalogue data cleaning and enhancement, and create a single data set of natural history specimens for data mobilisation. In terms of engagement activities, the NHM team will supervise community fellowships. They will also lead publications on the mobilisation of data held in Sloane's copy of *Historia Plantarum* and cataloguing of Sloane's botanical collections. The team members in BM research topics to inform the technical team and work collaboratively on a trans-institutional case study. They will lead the research and preparation for the touring exhibition with a number of venues, including the Down County Museum. There will also be permanent display changes in BM. The team will supervise the community fellows and, where fellows are amenable to the idea, it is hoped that fellows may also present their work at outreach events.

Project Partners

British Library	- To advise, including review of and input into project documentation, quarterly meetings, and active engagement/facilitation with Library colleagues and others outside the project. - General curatorial and project support from teams within Western Heritage Collections responsible for managing Sloane collections. - Provision of existing manuscript catalogue data from the Explore Archives and Manuscripts catalogue. - Provision of catalogue records for printed books held by the British Library, derived and enhanced from the Sloane Printed Books catalogue, and supplied as MARC XML.

Community Archives & Heritage Group	- Share experiences and to explore how the different solutions developed by the project could address similar challenges and inform the future strategy and work of CAHG. - To encourage members of Community Archives to actively engage in the project's participatory research sessions and attend meetings where the learning from the research will be shared. - Advise on the project's dissemination activities aimed at Community Archives and potentially co-hosting a community-focused conference or workshop. - Advertise the Fellowships over CAHG networks and potentially give high-level feedback on Fellowship applications that fall in their remit.
Collecting the West (Australian Research Council funded project)	- Committed to meeting with the Sloane Lab team for at least one 2-3 hour meeting a year for three years for the purpose of knowledge exchange and with the potential to develop more connections with the Sloane Lab team and their partners in Australia.
Down County Museum (N. Ireland)	- Host a travelling exhibition of key artefacts relating to Hans Sloane to help local people re-engage with the diverse aspects of Hans Sloane's legacy. It will be used as a touchstone to draw attention to the Sloane Lab project. - The Museum will use the Lab as a way to encourage its existing and new audiences to explore various themes relating to the lifetime and context of Sloane that are relevant to our lives today, including natural themes such as biodiversity, and cultural themes such as the legacy of colonialism and the slave trade. - A loan exhibition on Hans Sloane will enable the museum to explore all aspects of museology, using both the digital Sloane Lab and its own collections to stimulate interest, debate and audience participation.

University of Oxford, Department of Plant Sciences	• Knowledge exchange on sharing digital collections, using digital tools to support the digitisation of collections information, and learning how better support different people to use the digital collections.
University of Oxford, Gardens, Libraries and Museums	
Royal Botanic Garden Edinburgh	• To explore how the different solutions developed by the project could address similar challenges and inform future work and strategy.
National Museum of Scotland	
Historic Environment Scotland	• Contribution of staff time and potential for integration with other initiatives.
National Galleries of Scotland	
National Library of Scotland	

Staffing structure

Professor Julianne Nyhan, **Principal Investigator** (Professor of Digital Humanities, UCL and Professor of Humanities Data Science and Methodology, TU Darmstadt), is the principal investigator and intellectual lead of the project. Nyhan has overall responsibility for the research, technical and participatory direction of the project and for the line management of UCL-based PDRAs. Nyhan maintains a bird's eye view of the project, its progress and direction overall, tracking research execution and trajectory, including identifying required alignments and readjustments in work package outputs and looking ahead to ensure the preparations for each sequential stage of the project are in place in good time. **Nyhan and all Co-Investigators** also contribute to research output dissemination and delivery, including presenting the project at high-profile international events.

Dr Andrew Flinn, **Co-Investigator** (Reader in Archival Studies and Oral History, UCL), is the Deputy Principal Investigator of the project. He supervises the work of the Participatory Consultant (Alda Terracciano) and co-leads the project's participatory strategy, implementation and evaluation with Dr Nina Pearlman, working closely with Nyhan. Dr Flinn also opens pathways to community archives and heritage groups.

Dr Andreas Vlachidis, Co-Investigator (Assistant Professor in Information Science, UCL), is the technical lead of the project. He supervises the PDRF's implementation of technical architecture and leads the research of the methods for integrating the dispersed Sloane collection under a unified Knowledge Base. He also contributes to the work of the participatory design, liaising with the relevant team member in an iterative process that seeks to incorporate input into the design and development of the knowledge base and other connected dissemination of data retrieval and visualisation.

Dr Nina Pearlman, Co-Investigator (Head of UCL Art Collections, UCL), supervises the work of the Participatory PDRA (Marco Humbel) and co-leads the project's participatory strategy, implementation and evaluation with Dr Andrew Flinn, working closely with Nyhan. Dr Pearlman also opens pathways to dispersed and living collections.

Dr Mark Carine, Co-Investigator (Principal Curator in Charge, Algae, Fungi, and Plants Division, Natural History Museum), is a herbarium curator responsible for Sloane's botanical collections. He supervises the NHM PDRA (Dr Victoria Pickering) and oversees the mobilisation of the NHM data with Sloane's taxonomic index to the herbarium as well as case study research.

Dr Jeremy Hill, Co-Investigator (Head of Research, The Directorate, The British Museum), supervises the BM PDRF (Dr Alicia Hughes), Programme Coordinator (hired in year 2) and BM Consultant (Dr Kim Sloane), is responsible for coordinating the overall strategy and implementation of the BM touring exhibition and for communications regarding data exchange between the British Museum, UCL, Natural History Museum, British Library and other partners as relevant.

Dr Sushma Jansari, Co-Investigator (Tabo Foundation Curator: South Asia, The British Museum), will be assisted by the BM PDRF (Dr Alicia Hughes) and lead changes to the BM permanent displays.

Dr Daniele Metilli, Research Fellow in Advanced Data Architectures for Digital Humanities (UCL), is responsible for the Sloane Lab metadata mapping, semantic data modelling, knowledge base design and development, including ontologies, linked data and graph databases.

Dr Foteini Valeonti, Research Fellow in Systems Development and Research (UCL), is responsible for executing computational components and interfaces, enabling programmatic access to digital collections, data aggregation and federation and front-end (application) development.

Dr Jawad Sadek, **Senior Research Fellow in Artificial Intelligence and Natural Language Processing** (UCL), undertakes research on NLP and AI to implement solutions for information extraction, automatic handwritten text recognition and semantic enrichment of museum data and cultural heritage collections.

Marco Humbel, Research Fellow in Participatory Research and Collections as Data (UCL), is responsible for contributing to participatory research and design; for facilitating a series of participatory events with heritage organisations, and planning community fellowships.

Dr Victoria Pickering, Research Fellow (Natural History Museum), is responsible for mobilising Sloane's taxonomic index to the NHM's Sloane herbarium; executing machine actionable versions of Sloane catalogues, case study research, as well as supporting the community fellows.

Dr Alicia Hughes, Project Curator (The British Museum), is engaged in mobilising data on Sir Hans Sloane's collections held at the British Museum, as well as researching case studies on Sloane's collections as well as supporting the community fellows.

Dr. Kim Sloan, Historical Consultant (The British Museum), is an acknowledged expert on Sloane and PI on previous AHRC and Leverhulme-funded projects on Sloane's collection.

Dr Alda Terracciano, Community Consultant (UCL), provides consultancy on the elaboration of participatory methodologies and instruments for working with diverse communities and for the planning and delivery of engagement events.

Programme Coordinator (Post commences in Year 2, The British Museum) will develop and deliver the touring exhibition programme.

Hanna James, Project Administrator (Finance and Research Manager, DIS UCL), serves as the first point of contact for all stakeholders, and coordinates all participatory events and communication channels. She also assists with reporting to AHRC and documentation of findings of participatory events.

Overall programme

	Year	2021			2022												2023												2024												Lead*:
	Calendar Month	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	
	Activities /Deliverables	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	M34	M35	M36	Post-project period			
WP 1 - Participatory and Engagement activities																																										
1.1	Planning and preparation																																									Part
1.2	Workshops x 10: Query Service & KB											x 3								x 3			x 3			x 1																Part
1.3	One-to-one sessions & testing/demo sessions: data modelling											x 5			x 4			x 2				x 1										x 1										Part
1.4	Workshop x 2 & online tasks: Data Atlas												x1																			x1										Part
1.5	Workshop x 4: API & Web Application (Data discovery)												x1												x 2							x 1										Part
1.6	Community Fellowships (CF): Sloane Lab data model																	CF x 2							CF x 1																	Part
1.7	Community Fellowships (CF): API & Web Applications																		CF x 2									CF x 1														Part
1.8	Community Fellowships (CF): Final iteration																															CF x 4										All
WP 2 - Technological Actions																																										
2.1	Datasets Review and Acquisition																																									Tech
2.2	Setup and Install Technology Stack																																									Tech
2.3	Design and Refine Data Model																																									Tech
2.4	Prepare and Map Data																																									Tech
2.5	Enrich and Ingest Data																																									Tech
2.6	Deploy Sloane Lab																																									Tech
2.7	Disseminate SL Framework																																									Tech
2.8	Develop Tools and Web Services																																									Tech
2.9	Publish Linked Data																																									Tech
2.10	Documentation																																									Tech
WP 3 Collections Work Package																																										
3.1	NHM data review																																									CT
3.2	NHM Catalogue data cleaning and enhancement																																									CT
3.3	Vegetable Substances collection (data set mobilisation)																																									CT
3.4	Creation of a single data set of Nat Hist specimens in NHM																																									CT
3.5	Ray Historia Plantarum I & II – data mobilisation																																									CT
3.6	Ray Historia Plantarum III (appendices) - data mobilisation																																									CT
3.9	BM Research topics to inform Technical Team																																									CT
3.10	BM Research & Preparation for Exhibition with Downpatrick Museum																																									CT
3.11	Exhibition opens in Downpatrick																																									CT
3.12	Preparation to tour exhibition to other venues																																									CT
3.13	Research & Preparation for BM display changes																																									CT
3.14	BM Display changes open to public																																									CT
3.15	Trans-institutional case study																																									All
WP4	**Reporting and dissemination**																																									
4.1	AHRC/TaNC Written Reports																																									All
4.2	Report during meetings	x1		x1		x1	x1	x1	x1	x1	x1		x1	x2	x1	x1	x1		x2	x1																						All
4.3	Conference and Publications																																									All

*Lead: sub-team that lead activities and deliverables

Part = Participatory Team; Tech =Technical Team; CT = Collections Team (BM & NHM)

Alpha Release   Beta Release   Final Release

Annual Project Review   Annual Project Review

Events and consultations

Consultations

Dates	Consultations	No. of attendees / Respondents
14 Oct 2021	**Meeting with the Royal Society (RS)** to introduce the Sloane Lab and catch up with RS on the development of their digitisation programme. RS offered to do a survey of the Sloane material and was interested in joining the Advisory Board for the wider consultation and contribute to the project.	Attendees include RS Head of Library and Information Services, Digital Resources Manager, Sloane Lab (SL) PI and Historical Consultant of Sloane Lab.
16 Dec 2021	**A follow-up meeting with the Royal Society** to discuss Sloane data provided by RS.	Attendees include the Head of Library and Information Services, Digital Resources Manager, and Technical Lead of SL
14 Jan 2022	**Meeting with Down County Museum Curator and CEO/Founder of Sir Hans Sloane Centre** to discuss the scope of the partnership and the role of Down County Museum in the touring exhibition.	Attendees include Down County Museum Curator and CEO/Founder of Sir Hans Sloane Centre, BM Head of National Programmes, and SL team members
19 Jan 2022	**Initial Meeting with Project Partners and Advisory Board** The purpose of the meeting was to introduce Sloane Lab to the partner organisations and Advisory Board and to invite support from them. The team explained the background of the TaNC Programme, as well as the research approach of the project.	Total attendance 36, including 26 from partner organisations and the Sloane Lab advisory Board members who located internationally.

16 Feb 2022	**Meeting with BM & NHM colleagues** to introduce the Sloane Lab participatory and engagement plan and consultations.	Attendees include NHM UX Manager, BM Community Engagement Consultant, BM Head of National Programmes, and SL participatory team.
24 Mar 2022	**Meeting with British Library** to discuss how BL could engage with the Sloane Lab project, also the areas of involvement and consultations. The various BL curator also gave overviews on the Sloane materials housing in BL, and some technical practicalities.	Total attendance 10, including Head of Western Heritage Collections and Curators.
20 Apr 2022	**Meeting with Australian Project Partner, Collecting the West (CTW).** Following a brief overview of both the technical and participatory methods of SL, the group discussed the possible knowledge exchange areas; including overlaps of the two projects on how to handle and aggregate data and shared experience on community engagement.	Total attendance 11, including 3 from the CTW.
4 May 2022	**Year 1 Advisory Board meeting** Apart from reporting on Work Packages, the project team sought the expertise and experience of the Advisory board in connection with a number of questions, such as best practices in the integration of dispersed data sets and including the voices of small heritage organisations and non-hegemonic communities.	Total attendance 26, including 12 from the SL Advisory Board.
5-6 May 2022	**Jeremy Hill (Co-I) and BM Head of National Programmes N. Ireland visit**. He met with Down County Museum CEO and other staff for Sloane Lab travel exhibition planning.	Down County Museum and Sir Hans Sloane Centre staff members

| 20 May 2022 | Meeting with UCL Copyright Officer (Referred by Advisory Board member) for **consultation on CC Licence and IP of reusing Sloane Lab materials**. | Attendance: UCL Copyright Officer, PI and Project Administrator |

Conference presentations, Talks and Interviews

Date	Descriptions	Attendance
13 Oct 2021	Co-I Nina Pearlman presented at a **UCL Institute of Collections Webinar** and introduced Sloane Lab internally.	Attendance: less than 50 with local and professional reach.
18 Nov 2021	Co-I Mark Carine co-organised a workshop with Queen Mary University: ***Disciplinarity and the early-modern herbarium - A one-day research workshop.*** He referred to the Sloane Lab and underpinned the multidisciplinary nature of the project, an opportunity to engage researchers from different disciplines.	Attendance: less than 50 with local and professional reach.
03 Dec 2021	Julianne Nyhan Sloane Lab PI presented at ***Universität Bern seminar Digitality/Digital Culture(s).*** Presentation topic: Loss, absence and contestation in Digital Humanities through the lens of the collection of Hans Sloane (1660-1753)	Attendance: For PhD students and advanced Master students of the University of Bern. More than 50 with international and professional reach.
27 Jan 2022	Julianne Nyhan Chaired and presented at **The Society for the History of Collections** special meetings ***New Perspectives on Sloane*** to provide a platform for new research on Sloane.	Attendance: more than 50 with academics and students.
19 May 2022	**Communicating Colonial Legacies Workshop** A half day workshop organisedby TaNC and hosted by Sloane Lab in UCL	7 Sloane Lab members were present at the Workshop. Total attendance 32, including delegates from other Discovery Projects and TaNC/AHRC.

5-10 June 2022	NHM PDRA Victoria Pickering presented at the **Conference for the Society for the Preservation of Natural History Collections (SPNHC) Posters Session**. Abstract title: *Mobilising data in the Natural History Museum's historical botanical collection*	Attendance: The estimated number of participants is 450 with international reach.
5-8 Jul 2022	**Universeum 2022 (Belgium) - Poster section presentation** of conference paper Interrogating collections' contested and challenging past through the SLOANE (GLAM) LAB (J. Nyhan, A. Vlachidis, A. Flinn, N. Pearlman, M. Humbel, D. Metilli, J. Sadek, F. Valeonti, M. Carine, V. Pickering & A. Terracciano) Research Fellows, Daniele Metilli and Marco Humbel are **bursary awardees of the Universeum Pre-Conference Training Workshop (5-8 July)**. They co-presented the outcomes of the workshop.	Attendance: about 100 with international reach from the University museums and archives sector.
15 Jul 2022	Digital Humanities Summer School at Oxford - DHOxSS2022 (11 July 2022 – 15 July 2022). PI Julianne Nyhan presented keynote at Keble College Oxford. *"Thinking through the place of absence in the grand challenges of the Digital Humanities and Humanities Data Science".*	Attendance: more than 50 with students and DH scholars.

Research approach

Our research approach is pursued across three principal strands – which are interconnected but discussed separately below for the purposes of clarity – technical elaboration and aggregation; participatory engagement; and data-driven collection case studies.

1. Technical elaboration and aggregation

This work of technical elaboration and aggregation can be conceptualised as a process of retro-engineering the envisaged public-facing, open-access Sloane Lab, which will contain the data, tools and services that can support the cross-collection search and the digitally-augmented exploration and reuse of digital cultural heritage collections. Working from the vision to the actual implementation, the following technical steps have been identified as essential to delivering the Lab:

1. Installation of computational infrastructure and architecture to support the lab.
2. Data identification, extraction and enrichment.
3. Iterative elaboration of a unified schema and mapping of Sloane's structured and unstructured data to this schema.
4. Ongoing ingestion of data into the Knowledge Graph (triple store).
5. Development of query and visualisation dashboard, including tools and services that allow that data to be Findable, Cross-searchable, Accessible, Interoperable, and Reusable.
6. Iterative User-testing of interfaces and approaches.
7. Publishing and deployment.

As of July 2022, substantial progress has been made across stages 1-4. At the core of the Sloane Lab resides the Knowledge Base (KB), which provides a homogeneous data environment using formal semantics to allow data integration, semantic enrichment, and knowledge discovery across a disparate environment of resources. This is now operational. The most fundamental challenge of the KB will be the provision of a suitable semantic metadata schema for unifying the catalogues and enabling the Knowledge Graph to facilitate resourceful query, visualisation and fact-finding. This is now in its first iteration, as structured data from Sloane's catalogues are being mapped to our common conceptual schema. Research with established technologies such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), Linked Data, RESTful, and other API-based services, is ongoing in anticipation of the upcoming dynamic harvesting of structured data from the Natural History Museum, British Library and British Museum. So too, the scoping study to guide the selection of which digitally imaged handwritten catalogues will be transcribed using trained Handwritten Text Recognition models and encoded in TEI-XML assisted by automated Natural Language Processing (NLP) techniques has now been completed. Furthermore, persistent identification of the ingested data will be introduced in the process of resolving unique identifiers for the purposes of cross-reference and alignment and publication within the Linked Data cloud.
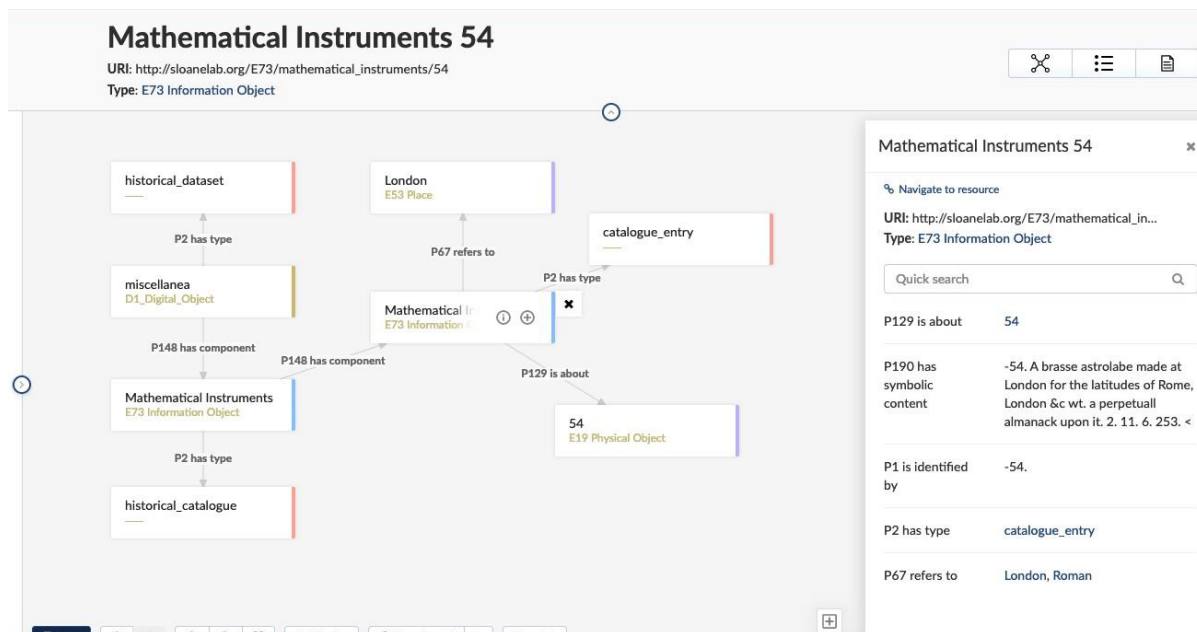
Figure 1: A screenshot of knowledge base alpha, with the corresponding mathematical instruments modelled as an E73 Information object.

Regarding technical step 2, it should be noted that the Sloane Lab Data Atlas is nearing completion. This is a substantial piece of work that has brought together previously dispersed information about the catalogues and datasets that currently exist, that are relevant to Sloane and which could be brought together in the Sloane Lab. To the best of our knowledge, this is the first time that the data-landscape of the foundational collection of the UK has been aggregated in print in this way. The Data Atlas will be of interest to a wide range of people, including those who are interested not only in the actual history of museum collections in the UK but also those who are also interested in the 'collections as data turn', including the new questions that we can and cannot ask when we treat culture as data. We expect this to be a major output of the project that has wide applicability and strong methodological relevance to TaNC more broadly.

An interface and visual identity for the Sloane Lab is far advanced. This interface has been developed to ensure that the website people will encounter when they visit sloanelab.org is memorable, easy to use, can deliver subjective satisfaction and that it can be used by the widest range of people possible, irrespective of any disability that a user may have (in otherwords, it is being designed to meet the highest standards of accessibility and usability).

2. Participatory Engagement

This project does not view the computational research of collections as separate from public engagement and impact. Accordingly, this strand of the project asks questions of the interface between collections as data, co-design and public engagement: how can we involve members of the public, researchers, curators and heritage organisations in the machine-actionable contextualisation of individual objects and collections, enabling community-sourced knowledge to enhance and even challenge existing Western-derived descriptions and conceptions? In answering such questions, we aim to unify historical and

present-day collections and records across taxonomy, geography, context and audience in a way that accommodates the interests of publics, expert communities and information organisations.

In recent months, considerable effort has been given to elaborating a robust plan for the planned participatory engagement strand, which undertakes up to 10 workshops, in addition to one-to-one sessions, testing session, use cases focus groups, and demo sessions with individuals and communities. The activities planned will be linked to the technical team's Sloane Lab Query Service and Knowledge Base tasks; the Sloane Lab data model; the Sloane Lab Data Atlas; and the Sloane Lab API and web application.

With activities on course to begin in September 2022, the planned participatory activities will result in new knowledge about what diverse people, expert communities and heritage institutions with different interests, concerns and knowledge want from a digital national collection. Closely aligned with the work and objectives of the technical team, the participatory sessions will not only involve users from a variety of professional, cultural, and geographical backgrounds that reflect the span of the collections; the sessions will also explore how the technical team can model data to support the interests of diverse communities. Looking to the future of the TaNC programme, the sessions are also expected to promote capacity building in the participating communities to increase their confidence in using a language shared with heritage practitioners as well as digital data.

The flagship of the participatory strand will be the 10 Community Fellows we will appoint through an open call. They will be given the funds and technical assistance they require to undertake a creative, research led or practice-based project on the digital national collection using the Sloane Lab. Along with a dedicated travelling exhibition, and changes to the British Museum's Enlightenment Gallery, we will also support other memory institutions in the UK and internationally by releasing our code, tools and recommendations. Following ongoing consultation, we will develop tailored demonstrators and recommendations that can facilitate participation in the digital national collection. Work is currently ongoing toscope the legislative aspects of the Community Fellows calls, the first of which we expect to be publicly advertised in January 2023.

Following a wide-ranging literature review and the drawing up of the relevant instruments, a series of interviews are now also in progress with representatives from heritage institutions. Interviews are being conducted to understand more about the factors that might constrain and/or further engagement between a wide range of actors and the digital national collection. The findings of this aspect of the project are expected to have the potential to inform the longer-term development of collections as a data strategy in the UK and provide clear guidance to TaNC about areas of investment that might be prioritised so as to support heritage organisations' engagement with the digitisation of collections over the longer term.

3. Data-driven, collection case studies

The work of the collections team (BM and NHM) will advance both the project's capacity to support data-driven collections as data research and exemplify the new kinds of knowledge that can be gained from this work. One strand of this is focusing particularly on data enrichment, to ensure that *Sloane's botanical collection online* can not only be made cross-searchable but fully integrated with the Sloane Lab. Working presently on three volumes of Ray's *Historia Plantarum*, in close collaboration with the technical team, the data this containsis now being enriched and linked to records of specimens held in the Natural History Museum's content management system (CMS) and images of those specimens where

they are available. Following this, work will focus on the three volumes of Sloane's *Vegetable Substances* (VS) catalogue, editing/cleaning of VS catalogue data (three volumes, 12,700+ entries), linking people and places in catalogue data to unique identifiers from the CMS and establishing protocol for serving digital images of catalogues from NHM and implement for VS catalogues.

Attention is now also turning to the planning of a collaborative case study on the Sloane collections' representation and acknowledgement of its colonial context, to be undertaken collaboratively by NHM, BM and UCL. Expanding the scope to work on collections held across institutions and on those catalogues not previously addressed by Enlightenment Architectures, the purpose of this trans-institutional case study is two-fold: it will act as an incubator and proving ground for the technologies and processes developed by this project, applying and evaluating them across the breadth of Sloane's catalogues and collection data, leading to further refinements of the project's automated and semi-automated workflows and techniques. It will also advance empirical and critical critiques of data-driven and quantitative approaches to identifying and visualising the colonial processes, infrastructures, agents and provenance of the national collection as embodied in Sloane's collection. Utilising the semantic modelling of entities that this project will unlock, we will, for the first time, be able to reassemble and map the stated geographical provenance of objects and specimens across the entirety of Sloane's collection; the role of trading companies and enterprises in the shipping of objects and individuals connected with the collection, creating new knowledge about Sloane's collections as a product of colonial activities. The case study will directly inform the planned physical and touring exhibition and has the potential to significantly impact understanding of the research capacity that can be created by making collections machine-readable.

A touring exhibition will form part of the established National Programme of the British Museum touring exhibitions. These are designed to best support the partner museum's ambitions and audiences. Those partner museums will shape the focus and nature of this proposed Sloane-focused exhibition. The exhibition will travel to at least four venues across the UK, including our project partners in Down County Museum.

Research results

The main outcome of the project will be a digital platform that can support research on digital national collections. This platform, the Sloane Lab, will comprise: a search interface (to the reunified present-day collections, historical catalogues and external resources), content (Sloane's historical catalogues as machine actionable datasets for reuse), tools for managing and researching the data (to support, e.g. network analysis, text analysis or AI-led interrogation of collections and records); documentation covering uses of the data (e.g. schemas, licensing details); and an engagement demonstrator that facilitates the participation of heritage institutions in the Sloane Lab (e.g. white paper and code-books).

An important research outcome of the project will be the methodologies and instruments that the project will devise not only for the technical aspects of pursuing collections as data research but also for working with diverse communities. Accordingly, the project will closely document not only its research outcomes that pertain to the new kinds of questions that individuals and communities might wish to ask of Sloane's collections, including the language they want to use; it will also document and make freely available the methodologies it is pursuing to support this work, including instruments for working with diverse communities, like interview questionnaires, workshop activities and so on.

Project outputs

- A digital platform, the Sloane Lab, which can support a new kind of data-driven engagement with the national collection.

- At least six co-authored articles will be published in international, peer-reviewed scholarly journals like the International Journal of Heritage Studies and Journal of Documentation in order to present the outcomes of the project to the international academic community.

- A series of outputs that document, and where possible, make executable, key aspects of the project's technical and methodological elaboration. Output in the final stages of the project, they will be output in a way that is appropriate to the eventual form each respective output takes, for example, policy-based recommendations in a white paper; best practice reports on, e.g. approaches to automating semantic and structural annotation of digital cultural heritage as a narrative report or dataset, whether released as a Jupyter notebook, via GitHub or another publicly assessable data repository.

- A Touring Exhibition for partner museums will be developed around a prominent object or objects in Sloane's collection chosen to best convey the work of the project and the stories about these objects that can only be told because of bringing the collection together.

- Changes to permanent exhibitions in the British Museum Enlightenment Gallery are also foreseen.

Early outputs:

PI J. Nyhan interviewed in: Jack, Andrew.

2021. 'Search for a Digital National Collection'. *Financial Times*, 22 November2021.

https://www.ft.com/content/1282bb6e- bbde-4449-8efc-42b43335f8f

Sloane Lab Newsletter - published four issues so far:

Issue 1: 22nd November 2021

Issue 2: 8th February 2022

Issue 3: 11th April 2022

Issue 4: 21st June 2022

The Sloane Lab Newsletters introduce the project team members and provide updates about the project to the key stakeholders and interested individuals. As of July 2022, it has 57 subscribers and an international reach (Australia, USA and Canada).

Cross Project Collaboration

Collaboration with other TaNC projects to date has centred around the technical aspects of the project, specifically on ontologies and vocabularies for data modelling. We have benefited from knowledge exchange with the UNPATH'D Waters project, who shared their *Towards the UNPATH'D Waters Ontology* draft report with us, which sets out "a core ontology to which UNPATH data providers will map their individual database schema, thereby ensuring interoperability between diverse resources" (UNPATH'D Waters 2021).

Sustainability and Infrastructure

Project management data, like progress reports, communication, schedule, notes, ethics and governance documentation, are being stored in UCL's Microsoft SharePoint and OneDrive for business server, managed by UCL Information Services Division (ISD) and backed-up daily. Our Sloane Lab SharePoint interface is available to all Sloane Lab team members based in UCL, the Natural History Museum and the British Museum.

Due to slow response rates, institutionally imposed security barriers to public facing URLs and scalability of resources, we have decided to move the project's infrastructure to AWS cloud services. Our current cloud services utilise Lightsail, EC2, Textract, Lambda, Route 53, Cognito, Backup, VPC, Certificate Manager, IAM, Event bridge. Two separate servers have been deployed. The WordPress server that hosts the project's website carries a specification of 512 MB RAM- 1 vCPU, 20 GB SSD. The development server that hosts the knowledge base carries the specification 2 GB RAM - 8vCPU - 200 GiB SSD - Operating System: Ubuntu 22.

Long-term data storage will be supported by UCL's Research Data Storage. The UCL Research Data service guarantees the preservation of research data for 10 years or more. The back-up is daily, and the service can be used during the lifetime of the project and after. The triple store data of the KB will be stored in these services, together with any additional research data. In addition, the technology stack of the VM and server configuration will be stored as a ready deployable Docker image.

It is also hoped that long-term storage will be supported by the BM and NHM, via their collections management systems, which guarantee preservation without time limit.

Contacts

Nyhan, Julianne - j.nyhan@ucl.ac.uk

Flinn, Andrew - a.flinn@ucl.ac.uk

Pearlman, Nina - n.pearlman@ucl.ac.uk

Vlachidis, Andreas - a.vlachidis@ucl.ac.uk

Jeremy Hill - JHill@britishmuseum.org

Mark Carine - m.carine@nhm.ac.uk

Hanna James – h.james@ucl.ac.uk

Annex

Figure 1: Sloane Lab Architecture

Source: Dr Andreas Vlachidis, UCL