Looking back to build future shared collections: reports from the Sloane Lab

Humbel, Marco

marco.humbel.17@ucl.ac.uk
Department of Information Studies UCL, London

Valeonti, Foteini

f.valeonti@ucl.ac.uk
Department of Information Studies UCL, London

Metilli, Daniele

d.metilli@ucl.ac.uk
Department of Information Studies UCL, London

Sadek, Jawad

j.sadek@ucl.ac.uk
Department of Information Studies UCL, London

Terracciano, Alda

a.terracciano@ucl.ac.uk
Department of Information Studies UCL, London

Pickering, Victoria

v.pickering@nhm.ac.uk
Natural History Museum, London

Hughes, Alicia

ahughes@britishmuseum.org
The British Museum, London

Vlachidis, Andreas

a.vlachidis@ucl.ac.uk
Department of Information Studies UCL, London

Pearlman, Nina

n.pearlman@ucl.ac.uk
UCL Art Collection, London

Flinn, Andrew

a.flinn@ucl.ac.uk
Department of Information Studies UCL, London

Carine, Mark

m.carine@nhm.ac.uk
Natural History Museum, London

Sloan, Kim

ksloan@britishmuseum.org
The British Museum, London

Nyhan, Julianne

j.nyhan@ucl.ac.uk
Department of Information Studies UCL, London; Technische Universität Darmstadt

Overview

Marco Humbel and Julianne Nyhan

The UK Arts and Humanities Research Council programme, Towards a National Collection (TaNC), seeks to "[…] take the first steps towards creating a unified virtual 'national collection' by dissolving barriers between different collections – opening UK heritage to the world" (TaNC, 2023) the published outputs of the TaNC programme to date have predominately (an exception being Pringle et al., 2022) addressed the potentials of technologies, such as IIIF, PIDs and Linked Open Data for creating aggregated collections (Padfield, 2020; Winters et al., 2022; Kotarski et al., 2022) and the implications of restrictive copyright frameworks for the reuse of collections (Wallace, 2022).

Yet, interoperability is also impacted by sociocultural factors (Borgman, 2015: 46–47). A digital collections' infrastructure, as much as any infrastructure, can be considered as "[…] a fundamentally relational concept; it emerges for people in practice, connected to activities and structures. It consists of both static and dynamic elements, each equally important to ensure a functional system" (Bowker et al., 2009: 99). In this panel, we will draw on our ongoing research to explore not only the legal and technical problems of creating a 'functional' digitally unified heritage collection, we aim to unpack, through the lens of the Sloane Lab project, the technical, historical and sociocultural elements that shape collections as data research and infrastructures in equally complex ways.

'The Sloane Lab: Looking back to build future shared collections', one of TaNC's five discovery projects, aims to reunify the currently dispersed collections of Sir Hans Sloane in a wider digital heritage infrastructure. Presenting new findings from this research, this panel will respond to the following questions:

Q1: What is the role of historical museum records in the task of reunifying siloed collections?

Q2: Which needs and priorities are to be considered for collaborating across the sector so that heritage institutions of different sizes can participate in unified digital heritage infrastructures?

Q3: How do we support the ability to search collections as data in innovative and deeply contextual ways?

Q4: How can we model and integrate disparate sets of historical and contemporary records from heritage institutions to facilitate interoperable access to digital collections as data?

Q5: How can heritage institutions benefit from cloud services to design a flexible and scalable technical infrastructure for digitising, releasing and interlinking their collections?

Q6: What methodologies do we need to employ to facilitate the use and interpretation of aggregated collections?

In this panel we offer a multifaceted and interdisciplinary perspective on the development of the Sloane Lab. We contextualise it both within a wider, nationally-funded drive to make the UK's cultural heritage, and information about it, computationally trac-

table and the more internationally-positioned work of 'Collections as Data'. Crucially, we also report on the Sloane Lab's participatory-design methodology which offers not only point of reflection on "[…] what kind of social values and ways of thinking and working are embedded in planned infrastructures […]", but also aims to "[…] contribute to the reconfiguration of the global representation of digital knowledge" (Pawlicka-Deger, 2021: 540).

Rationale for using heritage collections as a microcosm of the challenges of digital infrastructure integration (Q1)

Julianne Nyhan, Andrew Flinn and Nina Pearlman

How do historical museum records expose the complexity of digitally reunifying previously siloed collections in digital infrastructures? This presentation will emphasise the presentness of the past in collections as data research, and the complex dynamics that heritage collections pose to digital humanities. The extensive collection of Sir Hans Sloane (1660-1753), which is considered to be one of the founding collections of the UK's national collections attests to this. Assembled from c.1680s onwards, by the time of his death the Sloane collection was vast in scale and comprised tens of thousands of books and manuscripts, coins and medals, thousands of ethnographic objects and antiquities, hundreds of paintings and hundreds of thousands natural history specimens. This scale of collecting was financed by Sloane's earnings as a physician, profits from the enslavement of African people and the Transatlantic Slave Trade via investments in the Royal African Company, and dividends from his wife's plantations in Jamaica. The collection's assembly also depended on a vast network of individuals, up and down the social hierarchy and across different cultures, many of whom were part of a global trade of goods and people (Delbourgo, 2017: 202). The knowledge structures, such as naming, classification and cataloguing systems for different types of objects, that Sloane and others produced at this time through their collecting practices, often systemically silenced the voices, knowledge and methods of knowledge production of people and cultures who were 'collected'. As this presentation will explore, historical collections such as Sloane's demand a digital infrastructure that can contend with these histories, which continue to shape our present world namely: How can we make specialist users and members of the public more aware of the controversial and sometimes violent nature and histories of museum collections and of how digital platforms may perpetuate and even amplify this? What is the role of digital tools in foregrounding overlooked or hidden processes, like imperialism, colonialism, enslavement, loss and destruction, that have shaped the UK's national collections until now? We will present the Sloane Lab's initial answers to these questions.

Participating in aggregated collections – institutional perspectives (Q2)

Marco Humbel, Nina Pearlman, Andrew Flinn, Daniele Metilli, Andreas Vlachidis and Julianne Nyhan

A lack of financial resources, staff and expertise are often named as the most fundamental issues hampering heritage institutions in their ability to digitise (Pandey and Kumar, 2020: 30) and, by extension, to participate in digital infrastructures. Based on an extensive review of the literature we developed a questionnaire for semi-structured interviews to investigate the needs and priorities heritage institutions have regarding creating digitally unified collections. Semi-structured interviews are an appropriate instrument to explore digitisation practices in the heritage sector, as they have the potential to contextualise and extend information that is otherwise dispersed across a wide range of grey literature or not available to the public at all (Hauswedell et al., 2020: 140). In this paper we will report on our findings from the interview series which involved 18 key individuals responsible for collection management and digitisation in 10 heritage organisations and aggregators. We will synthesise the drivers and conditions such as strategic priorities, values, resources, expertise, and technical capabilities which present barriers for heritage organisations to participate in digital infrastructure projects. By doing so we will provide guidance to heritage infrastructure programs, such as TaNC, Europeana or the Digital Public Library of America on how to prioritise investments in heritage institutions to support their sustained participation in collections as data frameworks.

Mapping and visualising dispersed collections (Q3)

Foteini Valeonti, Andreas Vlachidis, Alicia Hughes, Victoria Pickering and Mark Carine

Searching collections that reside on disparate and scattered resources requires alignment and unification under a common application layer of data semantics for enabling innovative ways for cross-searching, contextual exploration and interrogation. How do we navigate a highly heterogeneous and divided environment of historic and contemporary catalogues? Building on the methodological foundations already laid by the Atlas of Digitised Newspapers and Metadata (Beals and Bell, 2020), Sloane Lab proposes the instrument of the 'Data Atlas' as crucial for answering this question. The Sloane Lab Data Atlas aims to accurately capture the project's data environment by mapping all data available relating to the Sloane Collections, the data atlas is the foundation that helps inform the process of data ingestion into the knowledge base. Beyond compiling and organising all available data sources into one comprehensive resource, the data atlas provides information for the different levels of availability (e.g. public, institution-only) and digitisation (e.g. imaged, fully transcribed). Critically, the Sloane Lab data atlas is expandable by design, providing templates for data collection, which are also used as an assessment tool, enabling consistent review, improvement and expansion. The benefits of the data atlas are numerous. Firstly, it provides a clear picture of the breadth and availability of all relevant resources and interrelations; secondly, by scoping the current state and format of those resources, the data atlas facilitates the design of appropriate data mapping and ingestion approaches; thirdly, it helps participant institutions identify significant data absences in their digitised collections. Most importantly, the data atlas provides a methodology for scoping out the data environment of projects dealing with disparate collections of thousands of objects and their digital surrogates, dispersed across institutions, allowing for a better understanding and decision-making in similar large-scale data projects.

Modelling and integrating catalogue records in an interoperable way – Questions and challenges (Q4)

Daniele Metilli, Andreas Vlachidis, Alicia Hughes, Victoria Pickering and Kim Sloan

The Sloane Lab aims to aggregate a multitude of catalogue records (both historic and current, from multiple disciplines) dispersed across the British Museum, Natural History Museum and British Library. The task of integrating these disparate records and facilitating interoperable access to them poses significant challenges. Sloane's historical catalogues are especially difficult to represent digitally, because the descriptions of the objects are often incomplete or inaccurate (Ortolja-Baird et al., 2019). Many catalogue entries lack information about the object's provenance, and even when this information is available, it is often vague. The fact that many objects are now missing and it is no longer possible to study them directly, exacerbates what Ortolja-Baird et al. call "the object problem" (2019: 21). When facing such challenges, it is crucial to adopt a critical digital heritage approach (Smyth et al., 2021), since the semantic representation of Sloane's historical catalogues may produce datasets that contain uncertainty and bia-

ses. To try to remedy such biases, we need to look for "data absences" in the records (Ortolja-Baird and Nyhan, 2022), and attempt to fill these gaps, or respond to them in an appropriate way. Moreover, Sloane's perspective on the collection is not easily reconciled with the perspectives of the institutions that hold the objects today. Every record embeds a certain worldview, and we should not always assume that there is a single correct interpretation. One potential answer to this challenge of 'multivocality' that we are investigating in the Sloane Lab is a data model that focuses more on the record than on the object, viewing records as different perspectives over the object.

Potentials for heritage institutions in adopting cloud infrastructures for digitising, releasing and interlinking their collections (Q5)

Jawad Sadek and Andreas Vlachidis

On-premises infrastructures impose technical challenges for system developers including restricted permissions, operational cost, and management overheads. This can reduce efficiency and cause performance issues. Cloud environments are becoming more popular in the industry and the academic community (Magoulas and Swoyer, 2020). Projects and organisations are adopting cloud-based solutions to respond rapidly to the demands of system agility and scalability. It is anticipated that heritage institutions can also benefit from cloud services to address the demands of complex system architectures. The technical and infrastructural requirements of the Sloane Lab demand scalable architectures, including but not limited to data mobilisation, knowledge base management and interrogation, data aggregation, serialisation and dissemination, and contribution to the Linked Data portal for enabling new ways to explore national collections. We aim to deliver open access applications that are constantly improved in a flexible manner and to develop a model for the sustainable preservation of the aggregated historical collections beyond the current period of funding. Our system is hosted on AWS where services and applications are packaged in Docker containers which are portable and can run on any infrastructure. The system architecture is structured in a way where applications and services are decoupled and deployed in separate layers. The Integration Layer hosts the data serialisation, aggregation, and semantic enrichment services while in the Presentation layer, we expose Sloane Lab's knowledge base for interrogation and interactive visualisations using the Metaphactory platform. At any given point the data workflow can be deconstructed and re-integrated with different platforms or technologies.

Participatory co-design: methodologies to facilitate the use and interpretation of aggregated collections (Q6)

Alda Terracciano, Marco Humbel, Daniele Metilli, Jawad Sadek, Foteini Valeonti, Alicia Hughes, Victoria Pickering, Andrew Flinn, Nina Pearlman, Andreas Vlachidis and Julianne Nyhan

The challenge for aggregating collections is not only based on a lack of technology or legal constraints. 'Soft factors', such as trust in technology, policies and incentives for participating in aggregated collections play an important role too. Moreover, developing a system for participatory modelling based on co-creation is a key priority to democratise the ways digital tools are created by shaping them around users' needs and aspirations. The Sloane Lab achieves this by "[…] opening up a space for intercultural exchange […]" (Terracciano et al., 2017), and reframing practices of participatory co-design of digital environments with communities of interest and heritage institutions. The method references Third Paradigm HCI (Harrison et al., 2011) and theories of co-creation of meaning that can facilitate a polyphonic, synchronic dialogue amongst different viewpoints, design elements and participants on issues related to the architecture of aggregators, their interactive elements and the historical significance and typology of material, "[…] to develop responses to research hypotheses and questions, to explore alternatives, and to reconfigure assemblies" (Bannon et al., 2018: 31). The discussion will cover selection of co-design participants, nature of activities planned around specific interests and abilities, and the incremental, iterative circular system for the participatory activities, meaning that the data created for and resulting from one activity feeds into the following one, both for the purpose of enriching information on a specific area, and to facilitate the flow of information across different users/co-design participants.

Acknowledgement