

Workshop Report
Legacies of Catalogue Descriptions: outputs and next steps
August 2021
cataloguelegacies.github.io

The project *Legacies of Catalogue Descriptions and Curatorial Voice: Opportunities for Digital Scholarship* is enabling transformational impacts in digital scholarship within cultural institutions by opening up new and important directions for computational, critical, and curatorial analysis of collection catalogues. Users and maintainers of collection catalogues know that extensive digital and digitised sets of curatorial descriptions from legacy catalogues are increasingly available to human and machine interaction. The *Legacies* project seeks to realise the potential of this data as valuable resources for cross-disciplinary research into curatorial practice, and for enhancing access to and analysis of collections at scale, as well as doing so in such a way that mitigates against the potential future harms of the legacy languages, structures, and knowledges they contain.



In July 2021 we held an ‘outputs and next steps’ workshop in order to disseminate work-in-progress outputs from the *Legacies* project and to work with community participants to develop next steps. Members of the project team spoke briefly on various aspects of the project’s work-in-progress research, before leading participatory breakout sessions that explored some of these themes in greater detail.

Over 40 participants from our target beneficiary communities (Museum and Information Studies Professionals, Digital Humanists, Cultural Heritage Technologists) in the UK and US attended the event, many of whom represent cultural institutions with significant legacy catalogue data and/or are aligned with policy networks related to cultural institutions.

Reports

James Baker ([Sussex](#)) reported on pilot research with the 1.1 million words written by the historian [Mary Dorothy George](#) between 1935 and 1954 that describe 12,552 Georgian satirical prints held at the British Museum. This research builds in two ways on previous research [funded by the British Academy](#): first, by analysing the ‘transmission’ of George’s style or “curatorial voice” from the British Museum to the [Lewis Walpole Library](#) in Farmington, Connecticut; and second, by being attentive to the harms this kind of legacy cataloguing - having migrated from text to database to online - can have in the world. In summary Baker argued that historical research into legacy cataloguing can usefully form the basis for reparative redescription and social justice work in cultural institutions.



Rossitza Atanassova ([British Library](#)) reported on the impacts a project of this nature has had on a large cultural institution responsible for complex and significant collection catalogues. Importantly the *Legacies* project aligns with a number of ongoing initiatives: [capacity building for digital scholarship](#), involvement in the the [Towards a National Collection scheme](#), and the Library’s [anti-racist action plan](#). Atanassova described how their involvement in the development of a significant project output - the training module '[Computational Analysis of Catalogue Data](#)' - and the decision to base that output around a selection of British Library catalogue data, enabled the project to demonstrate to colleagues at the Library how computational approaches to catalogue data can inform curatorial practice and help to accelerate institutional responses to contemporary challenges, for example with regards to the potential utility of computational analysis for appraising catalogue data and planning revisions to it.

Projects

Towards a National Collection



Image taken from page 137 of 'The Description of Britain, translated [by H. Hatcher] from Richard of Cirencester; with the original treatise *De Situ Britanniae*; and a commentary on the Itinerary [by T. Leman] Lat. and Eng' (London: J. Wright & Co., 1809).

British Library involvement in AHRC Programme Towards a National Collection

Published date: 31 December 2020

Living Knowledge blog

All our blogs Latest posts About this blog

Search this blog

Main

[Previous post](#) | [Next post](#) *

10 MARCH 2021

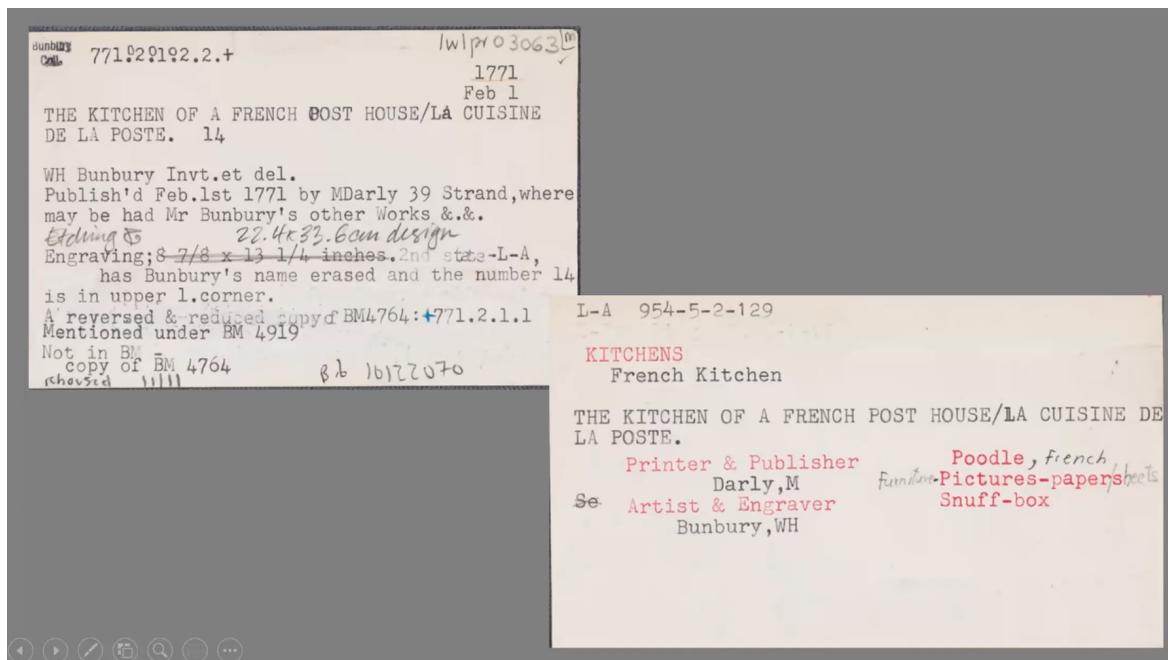
Towards an action plan on anti-racism



In July 2020, the British Library committed to create and implement an approved anti-racism action plan, sponsored by our Chief Librarian, Liz Jolly.

www.bl.uk

Cynthia Roman ([Lewis Walpole Library](#)) discussed ongoing research into the history of cataloguing at the Lewis Walpole Library that seeks to foreground the power of mid-twentieth century cataloguers and recognise their labour as steeped in historically-specific acts. Recounting the initial influence of Mary Dorothy George's work on early stages of cataloguing at the Library in the 1950s, Roman reflected on how interviews with past and present staff and gathering documents on cataloguing process has started to reveal both the adaptability of practice at the Library and the persistent influence - both directly and indirectly - of George.



Peter Leonard ([Yale University Library DH Laboratory](#)) introduced generative and experimental computational work that uses machine learning techniques - specifically, recurrent neural networks - to produce new texts based on historic catalogue data and new images based on print collections held

at the Lewis Walpole Library. Reflecting on these uncanny machine generated productions, Leonard encouraged us to consider what made these texts and images uncanny, and - when presented alongside ‘real’ catalogue data or images - what features made us think the machine output was real and which alerted us to the fact they were a fake.

Can you spot the fake?

1

Napoleon, much caricatured, stands before the Sphinx. He wears an academic cap and gown, and in his hand holds a scroll inscribed with the words 'Egyptology' and 'Science'. Behind him, a large pyramid with a golden cap; in front of him, a bronze statue. The inscription reads: 'Napoleon, Pharaoh of Science'.

2

Napoleon sits full face on a rock inscribed 'Usurped Power' and rising steeply from the sea. His eyes are covered by a green bandage between grotesque artificial eyes or disks from which rays slant down covetously to points in the sea. These are: (left) 'Great Britain', a small adjacent rock, 'West Indies', on the horizon, and (right) 'Malta', 'Egypt' [two pyramids], 'East Indies' on the horizon. He sits with elbows on knees and wears a huge cocked hat, and uniform with large epaulets and sword.

Finally, **Andrew Salway** ([Sussex](#)) [reported on](#) research to extend computational methods for analysing “curatorial voice” to the spatial and temporal transmission of catalogue data between catalogues. Focusing on the putative transmission of George’s catalogue descriptions to the Lewis Walpole Library, Salway described how transmission between catalogues can be detected by a combination of reading overviews of fields from digital catalogue records, corpus linguistic analysis of records for items with no equivalent at the British Museum, and deploying text distance metrics to directly compare catalogue records.

Text distance metrics

Edit-based metrics: count how many character-level changes would be needed to change one string into another.

Token-based metrics: treat the two strings as bags of words and measure how many words they have in common.

Sequence-based functions: return a value which gets higher as the two strings share more and longer sequences of words.

Our trials confirmed that a sequence-based function is best for ranking pairs of descriptions so that likely instances of transmission are at the top.

Breakouts

After the project reports, the event moved to three interactive breakout sessions. The first of these, led by **Baker**, introduced attendees to the training module ‘Computational Analysis of Catalogue Data’. The second, led by **Jake Kara** (Yale University Library DH Laboratory) demonstrated simCataloguer and encouraged attendees to think through how to read its machine outputs. In light of the project’s focus on foregrounding the situatedness of cataloguing labour, the third breakout, led by **Atanassova** explored the presentation of legacy descriptions in collection catalogues.

These sessions were lively, conversational, and productive. Upon reporting back to the group, overlaps and intersections between each breakout emerged. Participants emphasised the role the community could play in accelerating change, for example by investing in shared ownership of training modules like ‘Computational Analysis of Catalogue Data’ or by taking forward the ‘flagging’ of historic catalogue data. Participants noted that audit trails for legacy records are often noisy and incomplete, inhibiting both everyday curatorial work and creating caution over experimental reuses of catalogue data. Participants expressed uncertainty over how best to explain cataloguing conventions, genre, and silences to user communities.

Resources

As part of the project plan the *Legacies* team have been producing a series of resources intended to open up new directions for computational, critical and curatorial analysis of collection catalogues.

The first of these is the training module ‘Computational Analysis of Catalogue Data’, is accessible at cataloguelegacies.github.io/antconc.github.io/ and published under a [Creative Commons Attribution 4.0 International License](#). Published in February 2021, the module introduces people working in catalogue-related roles in galleries, libraries, archives, and museums to analysing catalogue data in AntConc, with a focus both on understanding how to use AntConc and how to use approaches from computational linguistics for the purposes of examining catalogue data. The module is built on [The Carpentries](#) pedagogical model, and as a result is ideal for self-directed learning, peer-learning orientated ‘[hack and yack](#)’ sessions, or formal Carpentries-style instruction.

In preparation for the ‘outputs and next steps’ workshop, we have began to produce worksheets to encourage and enable further critical reflection on catalogues’ legacies. In light of discussions at the workshop, these have been published on the *Legacies* site.

- ‘[Reading simCataloguer](#)’ presents examples of catalogue descriptions ‘written by’ a machine cataloguer - simGeorge - and asks you to consider a series of questions about the veracity of those descriptions, the worldviews that underpin them, and the harms they may cause. The worksheet should take no more than an hour to complete and is ideal for small-group work. The worksheet is published under a [Creative Commons Attribution 4.0 International License](#).
- ‘[Presenting Legacy Descriptions](#)’ presents examples of ‘flagging’ web content for its historic character (common, for example, on news websites), proofs-of-concept for comparable flagging of catalogues, and examples of records that might need flagging, and asks you to consider what we can learn from these examples with regards to improving the user experience of encounters with legacy catalogue data. The worksheet should take no more than an hour to complete and is ideal for small-group work. The worksheet is published under a [Creative Commons Attribution 4.0 International License](#).

Next Steps

A key objective of the *Legacies* project is to lay the groundwork for substantial, multi-sectoral, and data rich partnerships focused on researching legacy catalogues, exploring their impacts on the contemporary world, and taking action to mitigate their future harms. This workshop was one part of a suite of activities that have developed and concretised the project team's existing partnership, enlarged our network so as to develop shared agendas, and demonstrated to wider audiences the priorities for future research around catalogue data.

In early-2022 the *Legacies* team will hold a final partnership development event to take this agenda forward. We were delighted by the number of participants who expressed an interest in doing just that.

If you'd like to be involved, please get in touch with James Baker at j.w.baker@soton.ac.uk.