

Short Paper Abstract

The Great Unseen: Discoverability in Digitised Cultural Collections

Ellen Charlesworth, Durham University (0000-0003-0808-204X)

Paul Guhennec, École Polytechnique Fédérale de Lausanne (0000-0002-5490-5249)

Bianca Schor, University of Cambridge (0000-0001-5359-4656)

Increasingly, art galleries and museums have made their collections available to the public and researchers through online interfaces and databases. Ostensibly, this vast corpus of metadata and digital images improves access to researchers across the globe (Putnam 2016), yet as the number of digital resources grows, it has become unfeasible to exhaustively trawl through every potentially relevant item on the internet. In this context, search and ranking algorithms have come to mediate our access to online cultural resources (Kizhner et al. 2021). Our research explores to what extent UK museums have made their collections available online and how researchers' access item records. Looking at over 3,000 museum websites, we identified online collections, hosting infrastructures, and the types of search and browsing tools available. This paper will focus on the initial findings from British collections, which will be expanded to create a larger Europe-wide dataset over the coming months.

The preliminary results have clearly highlighted the ubiquity of free-text search, something that has long been of interest to galleries, libraries, archives, and museums. As Dan Lipcan summarised, "that a library user could, through a single search box replicating the Google experience, retrieve more than holdings in the print collection was undeniably attractive" (Lipcan 2021, 93). However, as search engines have evolved – integrating knowledge graph results, improved natural language processing, and AI co-pilots – the gulf between their capabilities and those of collections searches has grown. As these free-text search boxes are often now associated with natural language querying, not keywords, what assumptions may users be making about how these searches work?

The visual uniformity across interfaces in no way reflects a standardisation of the underlying search mechanisms. Despite this one design element being shared across multiple sites, some utilise keywords, others inexact string matching, and each prioritises different metadata fields. To users, it is unclear what dictates the ranking of results, as they are only shown the final homogenous list of outputs. The strictness of a database format imposes a form of standardised representation that is incompatible with even a simple algorithm's complexity (Panagiotidou et al. 2022). Collections searches do not usually utilise uninterpretable black-box algorithms – where it is unclear what contributed to a decision or prediction – however, even simple decision-making logic is not being clearly communicated to researchers. The realities of search are often unintuitive; in the prevalent 3rd party software Omeka, if a date is formatted as text to add a descriptive element, such as 'circa' or 'century', it will be sorted alphabetically. This kind of nuance is rarely made clear in the search interface itself or explained to users, potentially contributing to the biases already introduced to historical research by the digitisation process.

We argue that the solution to this hermeneutical challenge is to be found in the reconstruction of the chain of interventions – from the object itself, to its position as a search result. By making clear the algorithmically-constructed nature of the digital record and the choices that were made in its construction (Bachimont 2022), researchers are able to assess how this algorithmic mediation and the underlying digital infrastructure may introduce biases into our perception of the historical record.

While the largest museums in the UK have sophisticated interfaces, many without the resources to customise software are ill-equipped to communicate the nuances of search to users. This is especially true of smaller institutions, which heavily rely on off-the-shelf solutions for collections' management. By looking beyond major institutions – to those without the resources for cutting edge digitisation projects and custom software – we attempt to understand levels of digital access across different types of organisations in the UK. As such, we have been able identify a wide variety of approaches museums' have adopted towards explainability. During this ten-minute talk, we will pull examples of best practices from our broader dataset, asking how we might improve the transparency of search results and how the consequences of hidden algorithmic choices can be visualised.

In the context of continued digitisation and growing linked open data initiatives, we argue that addressing search transparency will become increasingly important and is key to researcher-user's interpretative process. Our works aims to capture the current state of the field, and using concrete examples of algorithmic transparency in practice, we take a step towards better explaining to researchers why certain material is presented to them and the intermediary roles of search and ranking algorithms.

Bibliography

- Bachimont, Bruno. 2022. 'Donner du sens aux données : les ruses du numérique'. *Interfaces numériques* 11 (2). <https://doi.org/10.25965/interfaces-numeriques.4838>.
- Kizhner, Inna, Melissa Terras, Maxim Rumyantsev, Valentina Khokhlova, Elisaveta Demeshkova, Ivan Rudov, and Julia Afanasieva. 2021. 'Digital Cultural Colonialism: Measuring Bias in Aggregated Digitized Content Held in Google Arts and Culture'. *Digital Scholarship in the Humanities* 36 (3): 607–40. <https://doi.org/10.1093/lc/fqaa055>.
- Lipcan, Dan. 2021. 'The Life of the (Third-) Pary (System)'. In *The New Art Museum Library*, edited by Amelia Nelson and Traci E. Timmons, 89–96. Lanham, MD, USA: Rowman and Littlefield.
- Moretti, Franco. 2000. 'The Slaughterhouse of Literature'. *Modern Language Quarterly* 61 (1): 207–28. <https://doi.org/10.1215/00267929-61-1-207>.
- Panagiotidou, Georgia, Houda Lamqaddam, Jeroen Poblome, Koenraad Brosens, Katrien Verbert, and Andrew Vande Moere. 2022. 'Communicating Uncertainty in Digital Humanities Visualization Research'. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. <https://doi.org/10.1109/TVCG.2022.3209436>.
- Putnam, Lara. 2016. 'The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast'. *The American Historical Review* 121 (2): 377–402. <https://doi.org/10.1093/ahr/121.2.377>.