

Our Heritage, Our Stories: Democratising the UK national collection

Hannaford, Ewan David

ewan.hannaford@glasgow.ac.uk
University of Glasgow, United Kingdom

Alexander, Marc

marc.alexander@glasgow.ac.uk
University of Glasgow, United Kingdom

Hughes, Lorna

lorna.hughes@glasgow.ac.uk
University of Glasgow, United Kingdom

Lewis, Rhiannon

rhiannon.lewis@glasgow.ac.uk
University of Glasgow, United Kingdom

Our Heritage, Our Stories (OHOS) is a collaboration between archivists, historians, linguists, and computer scientists from three UK institutions: the University of Glasgow, University of Manchester, and The National Archives. We aim to produce a fundamental shift in how digital content amongst communities across the UK is collected and curated; this poster outlines the project's goals, methods, and progress, and showcases examples of new humanities research stories that will be made possible by the project. *OHOS* is funded by the UK Arts and Humanities Research Council, as part of the Towards a National Collection research programme (project AH/W00321X/1), and runs until 2024.

With the continual diffusion of digital technologies, an ever-growing number of communities are independently gathering, creating, and curating digital materials relating to their own histories and cultures. These 'citizen histories' contain invaluable materials and perspectives that are not sustained by institutional collections, and which may challenge or change scholarly and societal perceptions across the humanities. This is especially true in the case of minority groups and other under-represented communities, whose cultural contributions have typically been side-lined by archival institutions in favour of 'treasured' material documenting mainstream populations, narratives, and linguistic varieties (Prescott / Hughes 2018).

As such, community-generated digital content (CGDC) is one of the UK's prime cultural assets. However, it is currently "critically endangered" (Digital Preservation Coalition 2021) due to technological and organisational barriers, and has proven resistant to traditional methods of linking and integration. The diversity of formats, storage practices, and archival standards through which CGDC is maintained by communities, as well as the heterogeneity of such content, poses a significant challenge to attempts to link and synthesise this data without eroding the complexity that is central to its significance. Consequently, prior attempts to integrate CGDC into larger collections have been limited.

As will be showcased in our presentation, *OHOS* is bringing together cutting-edge approaches from cultural heritage, humanities, and computer science to respond to this urgent challenge.

Through collaboratively recognising and ratifying the importance of CGDC across a broad range of disciplines, *OHOS* aims to enable the inclusion of a greater diversity of materials, peoples, and stories into the UK national collection, facilitating a wider and more accurate representation of UK heritage. Working with community groups across the UK to examine how they create and curate digital content, part of our project involves developing a model of best practice for creating and managing community-generated digital content. This innovative, co-designed post-custodial model will make generating and maintaining digital content easier and more streamlined, ensuring communities are able to digitally represent their content as completely and accessibly as possible. The work underpinning this model will also inform the production of research case studies, highlighting the new digital humanities research that will be enabled by the project.

OHOS is also developing a new resource that can ingest diverse CGDC into a centralised 'observatory', by which community collections from across the country may be explored and compared. Using Knowledge Graphs, this system will enable CGDC in a variety of formats and archival standards to be made interconnected and interoperable, while maintaining the informational richness of the original materials. These ingested materials will be further enhanced through the project's AI pipeline, drawing on established Natural Language Processing techniques, such as Named Entity Recognition, Relation Extraction, and Entity Linking, and enhancing these approaches for use with CGDC, such as through the integration of multilingual/multidialectal linguistic resources. With the implementation of additional computational approaches, such as vector-based semantic similarity metrics, the project will establish a fresh reservoir of community-generated materials for public and academic research and extend the work of community archives beyond their localities. These investigations will be facilitated by a "generous interface" (Whitelaw 2015) for the *OHOS* observatory, offering a wide-ranging suite of tools for exploring and 'remixing' CGDC. By affording novel means of connecting and comparing this previously isolated data, fresh avenues of humanities research will be revealed, and under-represented voices and communities will be brought to the forefront of our shared UK national collection.

Bibliography

Digital Preservation Coalition (2021): *The BitList 2021*. Available at: <https://www.dpconline.org/docs/miscellaneous/advocacy/wdpd/2521-bitlist2021/file> [04.10.2022].

Prescott, Andrew / Hughes, Lorna (2018): "Why do we digitize? The case for slow digitization", in: *Archive Journal* <https://www.archivejournal.net/essays/why-do-we-digitize-the-case-for-slow-digitization/> [27.04.2023].

Whitelaw, Mitchell (2015): "Generous interfaces for digital cultural collections", in: *Digital Humanities Quarterly* 9, 1.