

Investigations into a Global Digitisation Dataset

Stuart Lewis, Associate Director of Digital, National Library of Scotland

Dr Paul Gooding, Senior Lecturer in Information Studies, University of Glasgow

Mike Furlough, Executive Director, HathiTrust

Abstract

During 2019 a network of libraries from the UK and US were funded by the UK's Arts and Humanities Research Council under their Collaborations in Digital Scholarship programme to investigate the creation of a global dataset of all digitised texts¹ via prototyping and community engagement. This work was led by the University of Glasgow, with HathiTrust, the National Library of Scotland, the National Library of Wales, the British Library, and Research Libraries UK.

The main use cases for the dataset are:

1. Digital scholars seeking corpora of texts could search and compile links to items across many sources
2. Readers wishing to find a digitised text would be able to search efficiently across many sources
3. Libraries undertaking digitisation programmes would be able to avoid duplication in their own digitisation efforts

The project investigated these use cases and experimental work to understand the challenges of aggregating data from libraries with differing cataloguing standards and approaches. A trial dataset of over 17 million records was created and published openly² along with a final report³.

During the Covid-19 lockdown this concept was taken further with the development of a prototype search service 'OpenTexts.World'⁴ that contains over 8 million records.

This presentation will explore the findings of the project and discuss possible next steps in making the vision of a global dataset of all digitised texts a reality.

¹ Global Digitisation Dataset Project website: <https://gddnetwork.arts.gla.ac.uk/>

² Aggregated dataset of digitised texts from the GDD project, 2020. <https://doi.org/10.34812/fda4-5336>

³ Gooding, P. and Fulkerson, N. (2020) Towards a Global Dataset of Digitised Texts: Final Report of the Global Digitised Dataset Network. <http://eprints.gla.ac.uk/211898/>

⁴ OpenTexts.World: <https://opentexts.world/>