

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

2019 untersuchten mehrere Bibliotheken aus Großbritannien und den USA unter der Förderung des Britischen Arts and Humanities Research Council im Rahmen dessen Programms Collaborations in Digital Scholarship Möglichkeiten zur Erstellung eines globalen Datensatzes aller weltweit digitalisierten Texte¹ anhand von Prototypen und unter Mithilfe von externen Fachkollegen. Diese Kollaboration stand unter der Leitung der University of Glasgow, in Zusammenarbeit mit dem HathiTrust (USA), der National Library of Scotland, der National Library of Wales, der British Library, und des Verbandes Research Libraries UK.

Hauptendnutzer des Datensatzes:

1. Digitalwissenschaftler auf der Suche nach sammlungsüberschreitenden Datensätzen
2. Leser welche effizienten Schnellzugriff auf digitale Texte mehrerer Sammlungen benötigen
3. Bibliotheken die in ihrer Materialauswahl für Digitalisierungsprojekte Duplikation mit anderen Organisationen zu vermeiden suchen

Das Projekt untersuchte diese Fallstudien im Zusammenhang mit experimenteller Arbeit zur Ergründung von Herausforderungen die ein zusammensetzter Datensatz von Bibliotheken mit unterschiedlichen Katalogisierungsstandards mit sich bringt. Das Ergebnis ist sowohl ein Testdatensatz von mehr als 17 Millionen Daten welcher offen veröffentlicht wurde², als auch ein schriftlicher Bericht.³

Während des COVID-19 Lockdowns wurde dieses Konzept erweitert und durch eine Prototyp-Suchfunktion namens 'OpenTexts.World'⁴ ergänzt. Diese enthält mehr als 8 Millionen Dateneinträge.

Diese Präsentation wird die Ergebnisse des Projektes beleuchten und auf mögliche nächste Schritte eingehen um die Vision eines globalen Datensatzes aller digitalisierten Texte Realität werden zu lassen.

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/>