Datafication and reuse of the descriptions of the incunabula collection at the British Library

Atanassova, Rossitza

rossitza.atanassova@bl.uk
British Library, United Kingdom

The poster discusses my innovative research with the historical catalogues of the British Library incunabula collection, one of the largest and most important collections of pre-1500 published books (RLUK, 2022). By transforming the digitised legacy incunabula catalogues into data and performing computational analysis, I aim to create new knowledge about the incunabula collection and the Library's past curatorial and cataloguing practice, and to help update its online catalogues and improve the collection discovery. My research with the derived data from these catalogues supports the adoption by cultural heritage institutions of processes for the enrichment of descriptive metadata for print and digital collections that significantly benefits users, including digital humanists. This undertaking also highlights the importance of digital scholarship around legacy catalogues for the development of digital skills and research capabilities in staff, and the sharing of best practice with the wider library and archives practitioner community.

There is a broader context to my research which takes into account the recent initiatives in the Galleries, Libraries, Archives and Museums (GLAM) sector to critically re-examine and contextualise collections and legacy descriptions in order to be more inclusive and diverse, and evolve policy and practice. (ARLIS n.d.) The findings from my research also contribute to the better understanding of collection development policies and standards and the considerations around the ethical use of digital technologies with cultural data. (Ortolja-Baird & Nyhan 2022)

The poster outlines the digital research with *The Catalogue of books printed in the XVth Century now in the British Museum* that was published between 1908 and 2007. The thirteen volumes of the catalogue contain full descriptions of over 12,700 incunabula held at the British Library and is a highly used resource for research across many disciplines, including the history of printing, book trade and text studies. (The British Library n.d.) Firstly, I discuss the practitioner skills and processes required to derive structured data from the digitised images and reflect on some of the challenges of working with these resources and the suitability of the information for advanced computational analysis (The British Library 2022). Next, I present a case study of the reuse of the text data for the investigation of curatorial voices in the description of these collections for which I apply methodology from computational linguistics (Salway & Baker 2020). I also suggest different ways of presenting and sharing the datasets with the Digital Humanities and the wider research community, and invite a conversation about the opportunities for inter-sectoral and cross-disciplinary collaborations with the new digital resources.

In conclusion, I reflect on the strong collaborative element of my research, which has facilitated a dialogue between institutional staff: the current and former curators of the incunabula collection, staff responsible for collection metadata standards, metadata and digitisation analysts and digital curators. Simultaneously, I am working with a mentor from a Higher Education institution who brings expertise in the domains of history, cultural heritage and digital technologies and with whom I have recently collaborated on the development of training materials to perform computational analysis of legacy catalogue data. (Baker, Atanassova, Salway 2021)

Bibliography

Arts Libraries Society UK and Ireland (n.d.): *ARLIS Cataloguing and Classification Committee – Ethics series*. https://arlis.net/arlis-cataloguing-and-classification-committee-ethics-series/ [01.05.2023]

Baker, James / **Atanassova, Rossitza** / **Salway, Andrew** (2021): Computational Analysis of Catalogue Data https://cataloguelegacies.github.io/antconc.github.io/ [01.05.2023]

Ortolja-Baird, Alexandra / **Nyhan, Julianne** (2022): "Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive", in: *Digital Scholarship in the Humanities* 37, 3: 844-867 https://doi.org/10.1093/llc/fqab065 [01.05.2023]

Research Libraries UK (2022, May 4): *AHRC-RLUK Professional Practice Fellows announced 2022*: https://www.rluk.ac.uk/ahrc-rluk-professional-practice-fellows-announced/ [01.05.2023]

Salway, Andrew / **Baker, James** (2020): "*Investigating curatorial voice with corpus linguistic techniques*", in: *Museum and Society* 18, 2: 151-169 https://doi.org/10.29311/mas.v18i2.3175 [01.05.2023]

The British Library (n.d.): *Early Printed Books*. https://www.bl.uk/collection-guides/early-printed-books [01.05.2023]

The British Library (2022, November 29): *My AHRC-RLUK Professional Practice Fellowship: Four months on* https://blogs.bl.uk/digital-scholarship/2022/11/my-ahrc-rluk-professional-practice-fellowship-phase-one.html [01.05.2023]