British Library PhD Placement Scheme 2025/26 – Project Profile

Computational analysis of early printed book descriptions

Reference: 2025-WH-DSI

Supervisor(s)	Joint supervision: Karen Limper-Herz, Head of Printed Heritage Collections; Rossitza Atanassova, Digital Curator
Team(s) and department(s)	Printed Heritage Collections Digital Research Team
Start date/timeframe	Earliest start: June 2025 Latest end date: March 2026
Duration	**Three months** (or part-time equivalent)
Part time/Full time	☐ Full time only ☐ Part time only ☒ Full time or part time Minimum days/week if undertaken part time: two days a week
Location/Working arrangements	☐ Can only be offered on site at St Pancras ☐ Can be offered as fully remote placement ☒ Can be offered as hybrid placement (remote with some visits) One day per week or two days per month onsite

Context and nature of placement project

The British Library's collections of early printed books published in the 15th century (known as 'incunabula') comprises ca 23,000 volumes and is one of the most important and frequently used resources by researchers. The incunabula collection was catalogued in the 'Catalogue of books printed in the 15th century now at the British Museum [or Library]' (known as the BMC) over a period of 100 years between 1908 – 2007, and the detailed descriptions are a valuable source of information about the collection content, acquisition, transmission, previous ownership, and materiality.

As part of a previous digital scholarship project (https://app.transkribus.org/sites/BL-Incunabula) we used computational methods with the incunabula descriptions from BMC volumes 1-10 as means of improving our understanding of the collection's cataloguing history and curatorial practice. The PhD placement will help us to progress this work for two further volumes of the BMC and prepare more incunabula descriptions data for publication and analysis.

The Library's digital scholarship projects have demonstrated the value of providing access to our collections catalogues as data for research, public engagement, innovation and creativity. This work makes an important contribution to the Library's mission and especially to the Race Equality Action Plan as it creates and enhances metadata to remove barriers to discovery and improves the accessibility of our collection.

Tasks and outcomes

The primary focus of this placement will be to generate data with the incunabula descriptions published in volumes 11 (England) and 13 (Hebrew incunabula) of BMC. Under the guidance from staff in the Printed Heritage Collections and Digital Research teams, the student will use digital curation and data science tools to extract and prepare metadata for use in the Library's catalogues, conduct computational analysis and publish the research datasets on the Library's Shared Research Repository.

Typical tasks will include:

- Using Transkribus to train a layout model and output text (in XML format) from digitised images of volume 11 and volume 13 of the Incunabula catalogue
- Using Transkribus to adapt a Hebrew language model for the text in volume 13.
- Using and adapting existing code to extract individual catalogue entries from the XML output files from Transkribus
- Preparing a csv file with the individual catalogue entries
- Evaluating the metadata and preparing it for future ingest into the BL catalogues
- Preparing text files for each volume for analysis with the AntConc software
- Preparing the raw and processed files for publication on the research repository
- Using Named Entity Recognition and visualisation open source software with the data.

The main outcomes envisaged for this placement are:

- The creation of data that will be published in the BL research repository
- The creation of layout and language public models in Transkribus
- Corpus linguistics analysis of BMC volumes 11 and 13
- Blog posts about the placement project published in the Library's Digital Scholarship blog
- Staff talk as part of Researchers' Lunch programme

Training and development opportunities

PhD placement students are welcome to access a wide range of workshops, talks and training available at the Library. Supervisors will offer advice on which opportunities may be of particular relevance. Depending on availability, these wider training opportunities may include, for example, the Digital Scholarship Training Programme, Cultural Property Training, Research Roundtables, Business & IP Centre workshops and staff talks.

In addition, this particular PhD placement will provide the following opportunities:

- Introduction to the incunabula collection and cataloguing history
- Digitisation workflow resources and training
- Digital Scholarship Training Programme
- Research Services training events
- Inclusive Description Community of Practice resources and events

Required knowledge and skills
All applicants will be expected to demonstrate the following: • Ability to communicate effectively verbally and in writing, internally and externally • Working knowledge of Microsoft Office applications and online meeting tools • Ability to follow instructions and policies, in particular with regard to Health & Safety and the safe use of collections • Ability to work effectively with others while also completing tasks independently • Clear reasons for applying to a specific placement and good understanding of the project purpose In addition, this specific placement project requires the following: Essential: • Demonstrate an interest in Digital Humanities and/or an interest in early printed works • Willingness to learn and experiment with digital tools and methods • Interest in cultural heritage open data initiatives Desirable: • Experience of working with online catalogues, databases and digital repositories • Familiarity with digital methods • Some knowledge of python (or other programming languages)

Application deadline: **Friday 21 February 2025 (17.00 GMT)**

Eligibility criteria, funding information and details of how to apply are available on the British Library website: [Research collaboration - The British Library]