

Valid Validation – What's wrong with my digital file?

Reference: 2022-DP-FFV

Supervisor	Dr Peter May (Digital Preservation Technical Architect)
Team and Department	Digital Preservation Team
Start date/Timeframe	Earliest possible start date: June 2022 Latest possible end date: March 2023 (exact start date to be agreed with the supervisor)
Duration	3 months (or part-time equivalent)
Part time/Full time	Full time or part time For a part-time placement, the student would be required to spend a minimum of 3 days/week on the project.
Location and working arrangements	This placement can be offered on-site at London (St Pancras) or Boston Spa, as a hybrid placement (remote with some visits), or fully remote For remote working, the placement student would need a good broadband connection (to be discussed with the supervisor)

Context and nature of placement project

Digital content plays an increasingly important role in the British Library, with publishers delivering more content digitally and existing print collections being digitised. Assuring readers of the accuracy and authenticity of this content is vitally important. Digital preservation is a key endeavour in this pursuit, providing business and technical guidance to the issues affecting acquisition, long term storage, and access to digital content.

The Library is at the forefront of digital preservation research, with a dedicated digital preservation team who undertake research and engage with the international community – particularly through the Open Preservation Foundation and the Digital Preservation Coalition – to improve our understanding of practical approaches to digital preservation and guide their application.

The Digital Preservation Team's work includes technical research and development, collection-level assessments, risk analysis, and policy setting. Specifically, part of their role involves assessing the validity of digital content with respect to format specifications, understanding validation problems, and deriving technical guidance and policies on mitigating such problems. They use specialist format validation software tools (e.g. [JHOVE](#)) to check for compliance issues. These tools can identify a wide range of validation errors, though the varying impact of these errors on the long term renderability of digital content is not always clear and needs enhancing.

The placement student will work as part of the Digital Preservation Team on a deep-dive investigation of format validation problems affecting our digital collections, contributing knowledge back to the team and the wider community. The project will help the Library and the broader community to develop a better understanding of real-world format validation failures, and influence the preservation of digital collections in the Library's new repository system.

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 placement project will offer the following:

- Core Digital Preservation training, delivered either through our internally developed course, through 1-2-1 sessions with team experts, or through an external online course, depending on placement/course timings. Additional externally-provided digital preservation webinars may also be available (e.g. through Open Preservation Foundation or Digital Preservation Coalition).
- Specific training and guidance around file formats and validation tools will be provided throughout the placement by appropriate team members, and potentially through the Open Preservation Foundation (timing dependent). The exact nature will be dependent on the student's needs, but could cover topics such as: running/debugging validation tools, command line tool invocation, source code version control (Git), and Github wiki editing.
- Sharing of knowledge and results across the digital preservation community is important, and there will be opportunity to develop presentation skills through blog writing and giving webinars.
- Subject to availability at the time of the placement, internal Library courses, which may cover topics such as digitisation, metadata, content mining and cleaning data.

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:

- Computer programming experience, with a good awareness of how computers operate.
- Able to read and understand Java code (Python advantageous, but Java essential).
- Able to present and explain technical terms to a lay audience.
- Able to quickly read, understand and apply learnings from detailed technical documents.
- Able to focus and be detail oriented, but sufficiently grounded to see the bigger picture and explain findings in an accessible way.
- Knowledge of or interest in digital preservation.
- A detailed knowledge of format specifications is advantageous, but not essential.

This placement would suit a PhD student working in the field of Computer Science or those with a more technical interest in digital preservation.

Application deadline: **Friday 25 February 2022 (5pm)**

Eligibility criteria, funding information and details of how to apply are available on the British Library website: <https://www.bl.uk/research-collaboration/doctoral-research/british-library-phd-placement-scheme>