



LIBER 2026

Trondheim 1-3 July

Hackathons as Methodology Advancing Collections as Data Through International Participation

Ayla Karaman
Bodleian Libraries

Roman Kuhn
Staatsbibliothek zu Berlin



@libereurope.bsky.social



LIBER Europe



@libervideo



@libereurope



LIBER Europe

#LIBER2026



Project Overview



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Two events, one shared
approach to advancing
Collections as Data

Project Team

Ayla Karaman (Bodleian Libraries, University of Oxford)

Megan Gooch (Bodleian Libraries, University of Oxford)

Dorian Grosch (Staatsbibliothek zu Berlin)

Roman Kuhn (Staatsbibliothek zu Berlin)

Emma Lesburgueres (Staatsbibliothek zu Berlin)

John Woitkowitz (Staatsbibliothek zu Berlin)

Project funding

[GLAM-SPK Flexible Seedfunding](#) (Oxford In Berlin)



Code the Collection

Two-day open cultural data hackathon

4-5 March 2026

Centre for Digital Scholarship, Weston Library

Experiment and work with datasets provided by University of Oxford's
libraries and museums and the Berlin State Library



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



Core Questions

- What datasets are already available within our (wider) institutions?
- How accessible are they?
- How do people interact with them?
- How should we prepare/enhance them to make it easier for people to use them?

- Do we need to cater to different audiences providing different access points/formats/selections, etc.?
- What level of technical expertise can we assume from different users?
- What barriers currently prevent users from accessing or using the data effectively?
- What metadata and documentation are available to support data discovery and interpretation?
- How can we reduce the effort required to find and understand relevant information?

Institutional Backgrounds

Staatsbibliothek zu Berlin (Stabi)

- State Library (within the German federal system)
- Founded in 1661 (predecessor)
- Part of Stiftung Preußischer Kulturbesitz (Museums, Archives, Research Institutes)
- Services for broad audience, students and researchers
- no immediate affiliation with University

Bodleian Libraries

- University Library
- Founded 1602
- comprising 23 libraries in the University (excluding college libraries)
- Part of GLAM Oxford (Gardens, Libraries, Archives, Museums)
- Services for university members and researchers



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz



**CENTRE
FOR
DIGITAL
SCHOLARSHIP**



Institutional Backgrounds

Stabi Lab

central hub for the digital humanities and social sciences at the Stabi. The Lab facilitates access to data collections, the use of digital methods and tools, and data literacy in working with the Library's digital collections



HOME

DATEN

Aktuelles



Retrokonversion von Katalogkarten mittels Vision Language Models



Vergangenheit sichtbar machen: Der Editathon „Koloniale Spuren an der Spree“



Provenienz-Explorer

Das Stabi Lab ist der zentrale Hub für die digitalen Geistes- und Sozialwissenschaften der Staatsbibliothek zu Berlin. Das Lab vermittelt den Zugang zu Datenbeständen, den Einsatz digitaler Methoden und Werkzeuge sowie Datenkompetenz im Umgang mit den digitalen Sammlungen der Staatsbibliothek.

Als offener Raum für das Experimentieren mit Kulturdaten lädt das Stabi Lab ein, sich mit den Sammlungen und Daten der Staatsbibliothek kreativ auseinanderzusetzen. **Kontaktieren** Sie uns und wir beraten Sie gerne in Ihren Projekten.

Veranstaltungen

29.06. – Berlin Book Studies Network – Provenienzdaten an der Stabi

6.07. – Datencafé – MultiSlavDict: Historische slawische Wörterbücher gemeinsam durchsuchen und auswerten

9.07. – Blauer Salon – Werkstatttreffen zu RSE und digitalen Projekten

Daten



Type specimens



Bildannotationen des Datathons „Vom Polygon zum Dataset“



Thematische Datensets – Hackathon Coding Precarity

Centre for Digital Scholarship

physical and virtual hub, fostering innovation, collaboration, and knowledge-sharing in the realm of digital scholarship

About the Centre

The Centre for Digital Scholarship (CDS) is a leading and shaping discussions around digital scholarship at the University of Oxford. Housed within the Weston Library, our centre serves as both a physical and virtual hub, fostering innovation, collaboration, and knowledge-sharing in the realm of digital scholarship.

At the Centre for Digital Scholarship, we aim to:



Connect

Connect the Bodleian Libraries with broader national and international conversations, research, and advancements in digital scholarship.



Research

Research and innovate at the cutting edge of developments in digital scholarship



Lead

Lead and support research projects related to digital scholarship across the University of Oxford.



Support

Support the academic community, across all disciplines and levels of experience, from established digital researchers innovating in the field, to students getting their first taste of digital scholarship, either working

Hackathons as methodology

Why Hackathons?

- Participatory
- Experimental
- Cross-disciplinary



What can
hackathons teach us
about how people
actually use
collections as data?

Data, Events and Output

Our approach

- Oxford + Berlin
- Researchers, students, developers, creatives
- Open GLAM datasets (Oxford & SPK)

What participants did

- Built prototypes
- Explored datasets
- Created new use cases
- Combined and challenged existing data



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Datasets & APIs



Echoes of
Berlin: 78s

29.09.2025



Views of Berlin

06.10.2025



Lest we forget

07.10.2025



On the menu

01.10.2025



Digital
Bodleian
Collections

07.10.2025



Stereocards

24.09.2025



Secret State
Archives:
Descriptive



Ancient Coin
Data



Typographia
Sinica

24.09.2025



Key findings



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

CENTRE
FOR
DIGITAL
SCHOLARSHIP



How we documented the hackathons

Project learnings

Next steps



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz

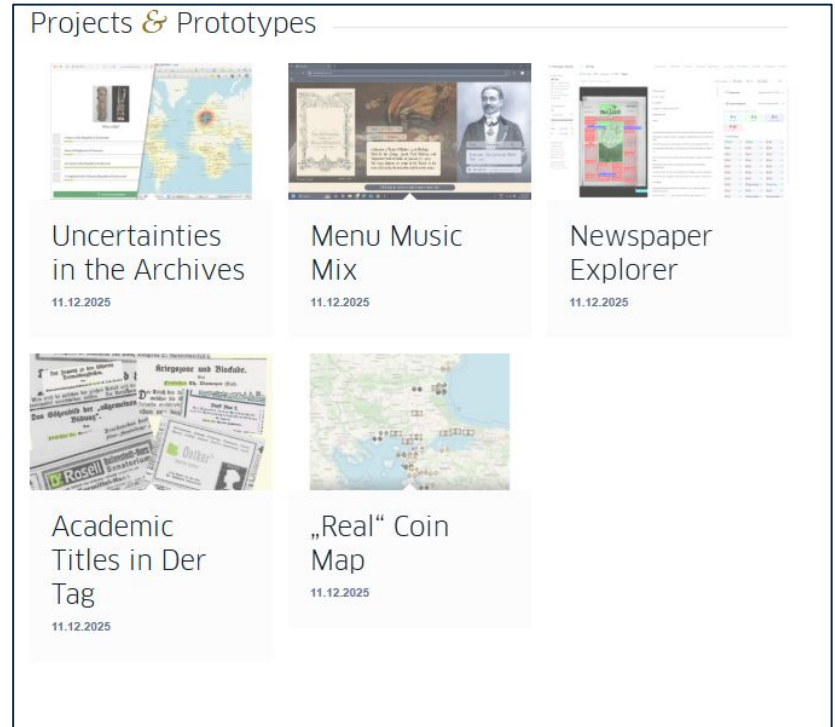


CENTRE
FOR
DIGITAL
SCHOLARSHIP



How we documented the hackathons

- Sharing the group projects
- Feedback
- Observation



Group projects: New & unexpected use cases

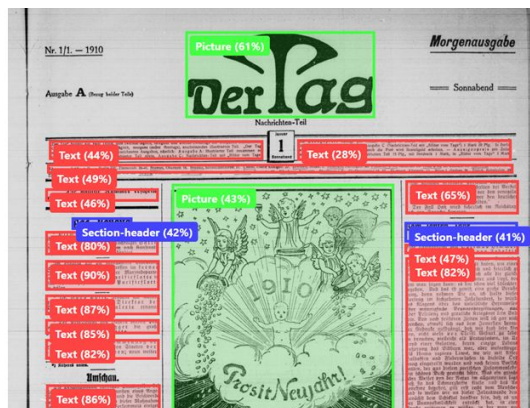
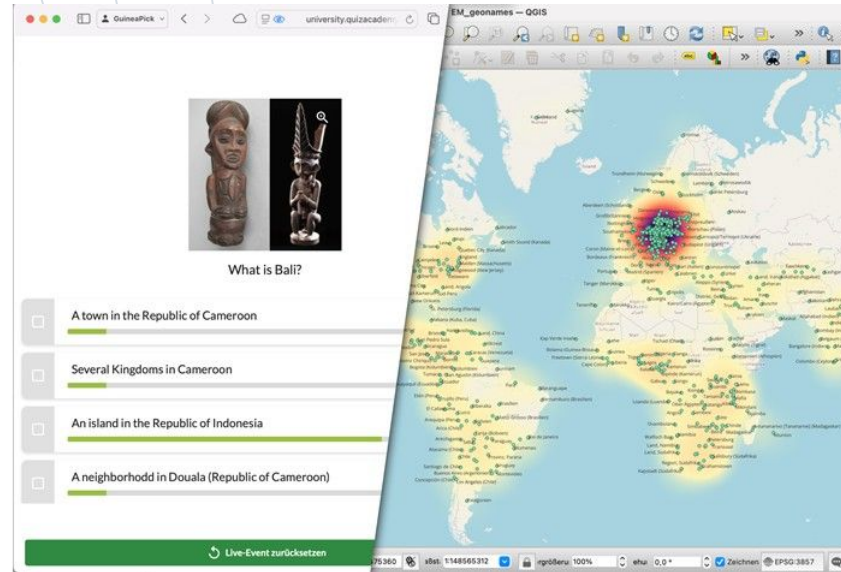


Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz

Stabi
Lab



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Feedback

We wanted to understand the value of GLAM datasets beyond the library and to understand the ways these datasets are used by researchers/students/readers.

We circulated the same feedback form for both hackathons, so we could compare the results.



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

CENTRE
FOR
DIGITAL
SCHOLARSHIP



Feedback

Background of surveyed participants

Berlin hackathon

52% of attendees provided feedback

Oxford hackathon

67% of attendees provided feedback



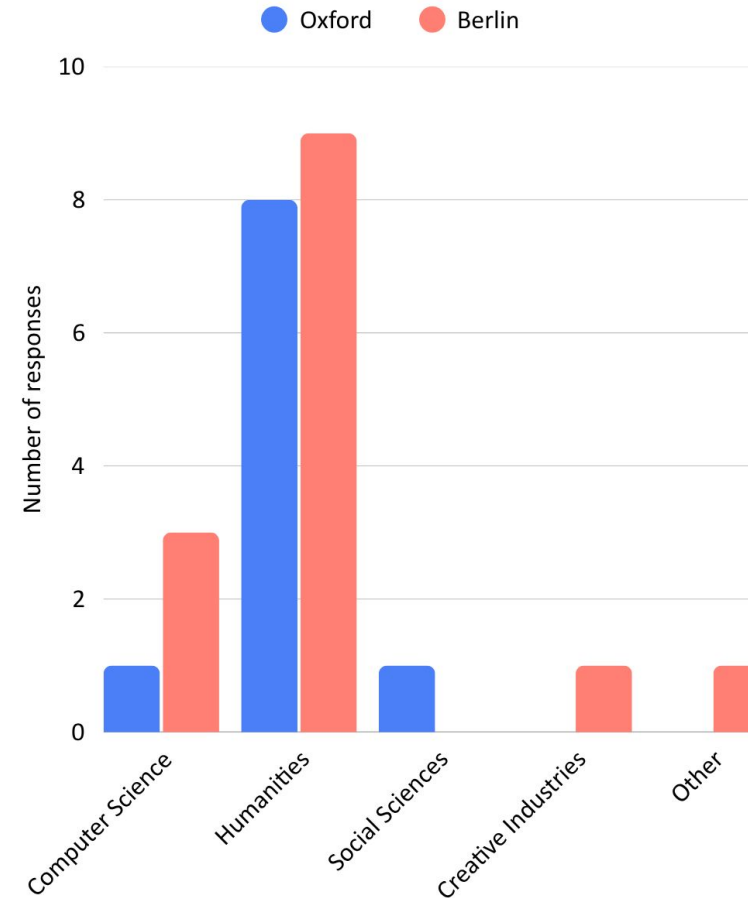
Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



How would you describe your
professional/academic/research background?



Feedback

What kind of documentation or support would have helped?

I think some preprocessing on the XMLs to extract plain text of the text blocks would have been quite helpful, though we managed to vibe-code a parser pretty quickly.

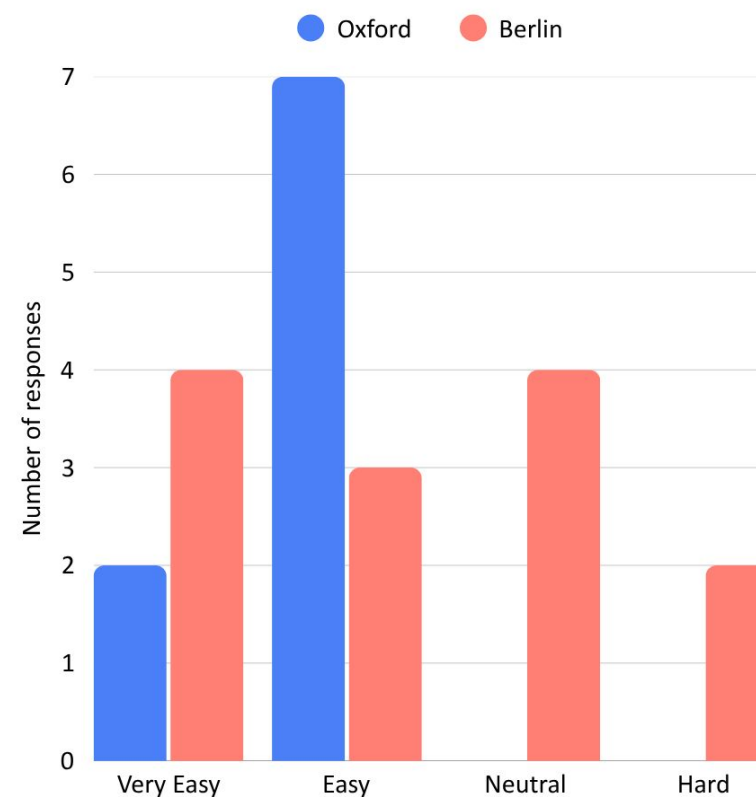
More information on how the dataset was constructed, its format etc.

A few workshops or even just resources on how to use or begin using different platforms for data analysis (github, python, IIIF, etc.)

A short guide introducing basic concepts would have been helpful, so as to be familiarised with the main ideas/tools beforehand

● Berlin feedback
● Oxford feedback

How easy was it to understand or access the data?



Base: 9 responses from Oxford, 13 responses from Berlin

Feedback

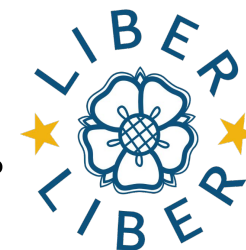
Did you feel your team had the right mix of skills?



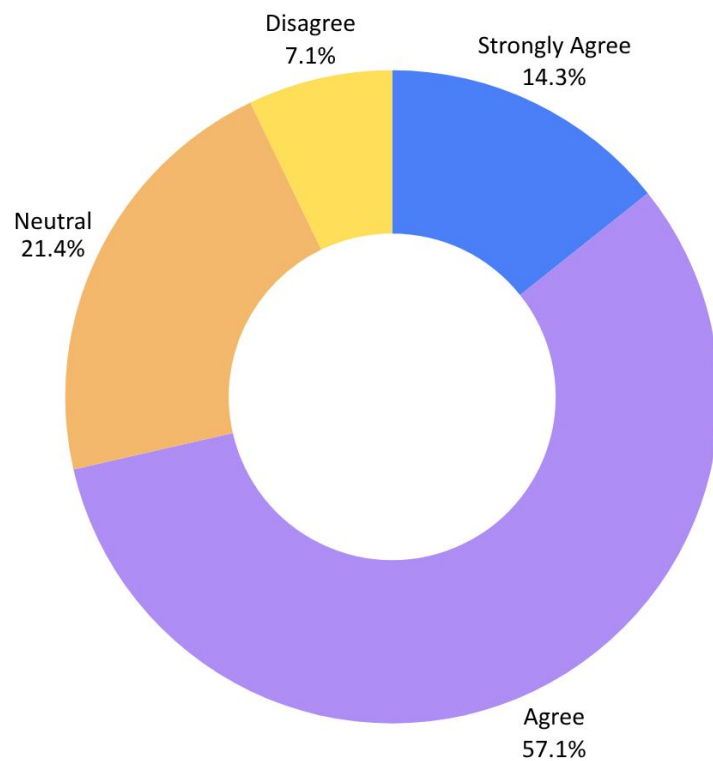
Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP

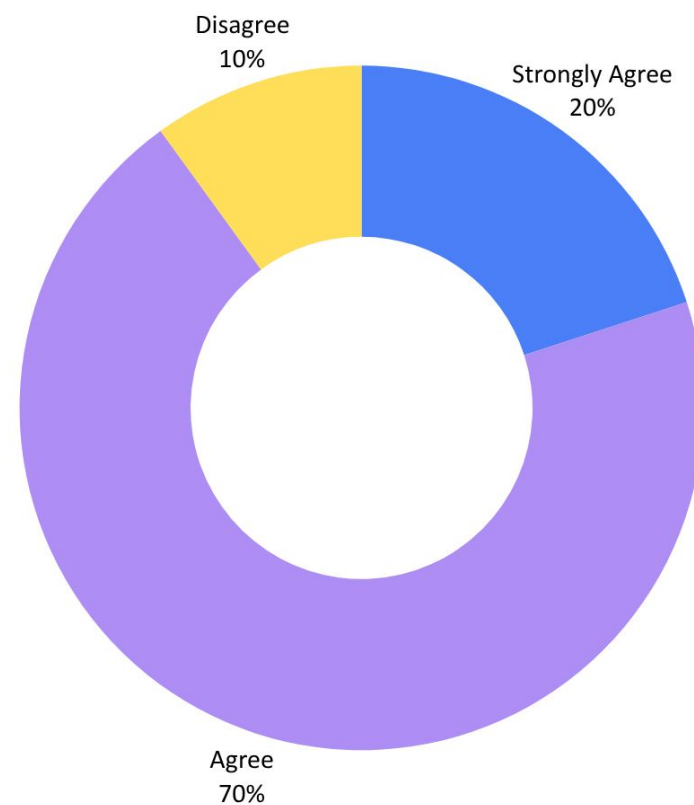


Berlin hackathon



Base: 14 responses

Oxford hackathon



Base: 10 responses

Feedback

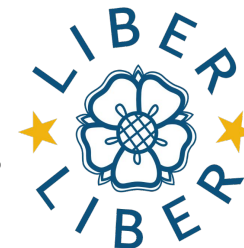
Did you learn a new tool or method during the hackathon?
If so, what tools did you learn?



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz

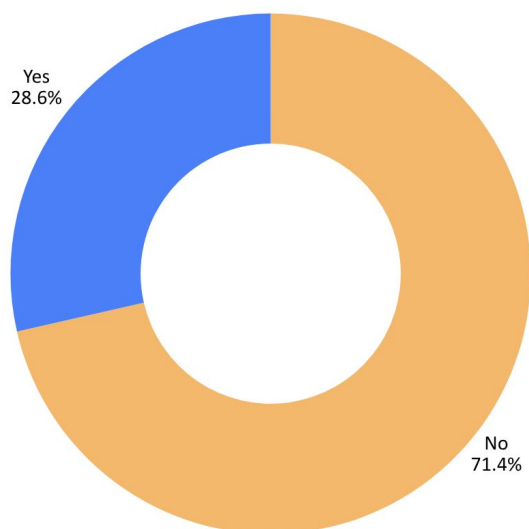


CENTRE
FOR
DIGITAL
SCHOLARSHIP



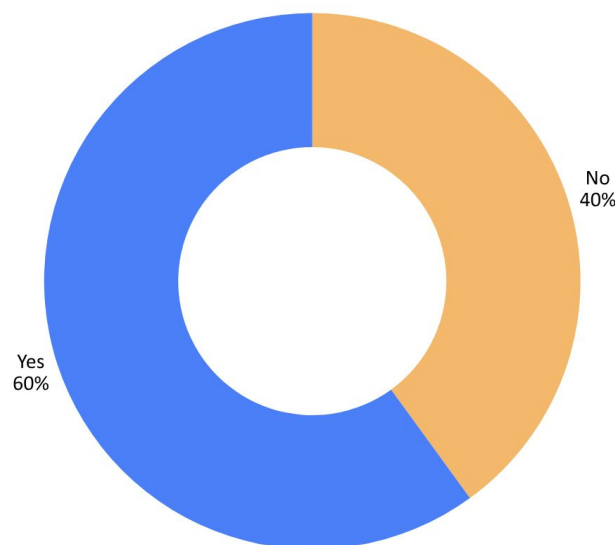
Tools/methods learnt
during the hackathons

Berlin hackathon



Base: 14 responses

Oxford hackathon



Base: 10 responses

mapping data
torchvision
transkribus
VS studio code
vibe coding
pytorch
geonames
GIS maps
dialect analysis

Observations



*Photo taken by A Karaman 7 Oct 2025,
Staatsbibliothek zu Berlin*



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

CENTRE
FOR
DIGITAL
SCHOLARSHIP



- **Number of datasets:** we offered a 25 datasets in total and 8 datasets were used. A few groups worked on the same datasets. This could be a mixture of interest and/or access ease within the limited time frame of the event.
- **Length of event:** Through verbal discussions, some groups would have preferred the event to be longer, but others (especially those travelling from further away) would have liked a one day option.
- **Technical abilities:** this varied and we did advertise the events as open to all. In some groups this did mean relying on those with more computational skills to produce the final digital product.

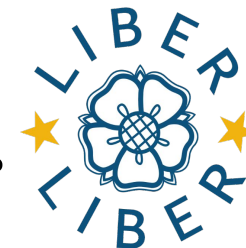
Observations



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



— CODE THE COLLECTION · OXFORD & BERLIN · 2020

Hackathon *Project* Proposals

Six proposed systems derived from 10 open cultural datasets spanning Oxford's libraries and Berlin's collections — each designed to solve a real discovery, access, or preservation problem.

10 DATASETS AVAILABLE | 6 PROPOSED PROJECTS

01 - Available Datasets

OXFORD · ASHMOLEAN

Ashmolean Museum Metadata

Antiquities, Western & Eastern Art, Coins, Casts
— broad multi-domain collection metadata.

[.CSV](#) [.JSON](#) [CC BY-NC](#)

OXFORD · ASHMOLEAN

Ancient Coin Data

~300,000 coins: Greek, Roman, Celtic, Byzantine, Islamic, Indian, Chinese coinages in XML.

[.XML](#) [CC BY-NC-SA](#)

OXFORD · BODLEIAN

IIIF Collections (Digital Bodleian)

Ephemera, Armenian Manuscripts, Literary Manuscripts — IIIF manifests with image API access.

[.JSON](#) [IIIF API](#)

Imran Asif, a Research Software Engineer for Digital Scholarship @ Oxford, created a project proposals page after observing the Oxford hackathon:

- Available datasets (selected datasets for the event)
- Problems to solve (data querying, discoverability issues, lack of geographic context, etc)
- Proposed systems (GIS & explorer maps, network visualisation, named entity explorer)

[Code the Collection — Project Proposals](#)

Project learnings

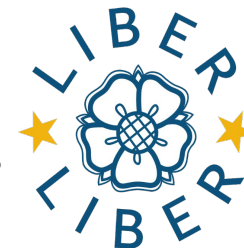
Value of experimentation



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Thank you so much! The MSc people had great fun to also be not so serious about their data practice.

We received positive feedback on creating a space where people can meet and exchange ideas, whilst creating innovative projects.

However, there were some comments noting the difficulties of learning new skills in this environment and relying on more tech-experienced attendees to produce the digital project.

It is a bit difficult to learn from others in this event. If you don't get explained how topic modelling works or knowledge graph, it's not likely you pick something up from these methods.

Project learnings

The benefits and limits of using hackathons to understand how people actually use collections data.



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

CENTRE
FOR
DIGITAL
SCHOLARSHIP



Benefits

- Discover in realtime how collections data is used and work with participants to navigate their projects.
- Participants can learn from each other and collaborate on exploring collections data in new ways.
- Learning about different user groups and how they access data and what support they need.
- An opportunity to collect qualitative feedback and find out what users need to better access and explore collections data.

Limitations

- Time-limited events mean participants might have chosen datasets that are easier to access.
- Not reflective of all library users accessing collections data.
- A few groups chose the same datasets, therefore feedback doesn't accurately represent all datasets presented.
- Learning new technical skills to access and explore the data can be limited when depending on technical members of the group.

Project learnings

Overview and next steps



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



- **Collaborative work informing strategic service development at both libraries**

Further understanding users needs for accessing data at scale and helps assess current collection data infrastructure.

- **Value of experimentation for future event/training programming**

Finding out what support, guidance and training is needed to encourage more users to explore the datasets.

- **Raising awareness of GLAM datasets**

Creating events such as these opens up access to collections data and produces new and unexpected digital projects and use cases.

Next steps



*Photo taken by A Karaman 7 Oct 2025,
Staatsbibliothek zu Berlin*



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

CENTRE
FOR
DIGITAL
SCHOLARSHIP



- **Publishing results and project learnings**

We want to share our learnings to help other libraries interested in delivering a similar initiatives. Using our results, other institutions could expand further on 'hackathons as methodology' and continue to understand how users interact with open cultural heritage data.

- **Opportunity to expand current library services**

The event has highlighted areas we could explore for training in libraries. Expanding technical skills workshops.

- **Collaborating on future projects and events**

Provide resources and guidance on delivering similar hackathons, opportunity to collaborate with new institutions.



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



CENTRE
FOR
DIGITAL
SCHOLARSHIP



Thank you!

Please feel free to reach out for more information
on the project.

Ayla Karaman
Bodleian Libraries

Roman Kuhn
Staatsbibliothek zu Berlin



@libereurope.bsky.social



LIBER Europe



@libervideo



@libereurope



LIBER Europe

#LIBER2026

