The Future of AI and Libraries: Opportunities and Challenges at the Bodleian

Lizzie Hatfield, Head of Office for Strategy & Delivery
and Andrew Bonnie, Chief of Digital Operations

BODLEIAN
LIBRARIES
UNIVERSITY OF OXFORD

What does the Bodleian do and why?

AI and library services

- Can AI be used for sensitivity reviews for archives?
- How accurate are AI tools in helping with our curatorial work?
- Will scholars need to access today's algorithmic data for future research?
- How is AI changing how our readers find books and speak to subject librarians?
- What Text and Data Mining or AI for bibliographic research tools should we provide?
- What AI skills do students and researchers need?

AI and library collections

- Who owns AI-created works? And with widespread web-scraping, how safe is our own IP?
- Does handwriting text recognition work in library workflows? How accurately?
- Can (or should) we turn off web scraping by AI companies?
- Are our system providers using AI – and do we want to use those AI functions?
- Can AI create structured data (e.g. MARC records) from unstructured data (e.g. full texts)?
- What does 'open' mean in the age of AI?
- What does AI-created records mean for trust researchers have in the Bodleian?

AI and operating a library

- What impact will AI have on our commercial income (e.g. image sales)?
- Could AI companies fund our digitisation?
- How do we talk about our AI work to our readers?
- How can I use AI for personal productivity or as a personal assistant?
- What AI skills and knowledge do our staff need?

Where do we begin?

- Make sure your data is clean and structured
- Don’t build your own LLM
- Work with likeminded partners
- Build confidence using TREs and tech – pilot!
- Agree architectural principles and ethics
- Consider sensitivities of working with big tech

Our areas of exploration

AI and library services

Will scholars need to access today's algorithmic data for future research?

What AI skills do students and researchers need?

AI and library collections

Who owns AI-created works? And with widespread web-scraping, how safe is our own IP?

Does handwriting text recognition work in library workflows? How accurately?

Can (or should) we turn off web scraping by AI companies?

Can AI create structured data (e.g. MARC records) from unstructured data (e.g. full texts)?

AI and operating a library

What AI skills and knowledge do our staff need?

Could AI companies fund our digitisation?

Bodleian-OpenAI pilot

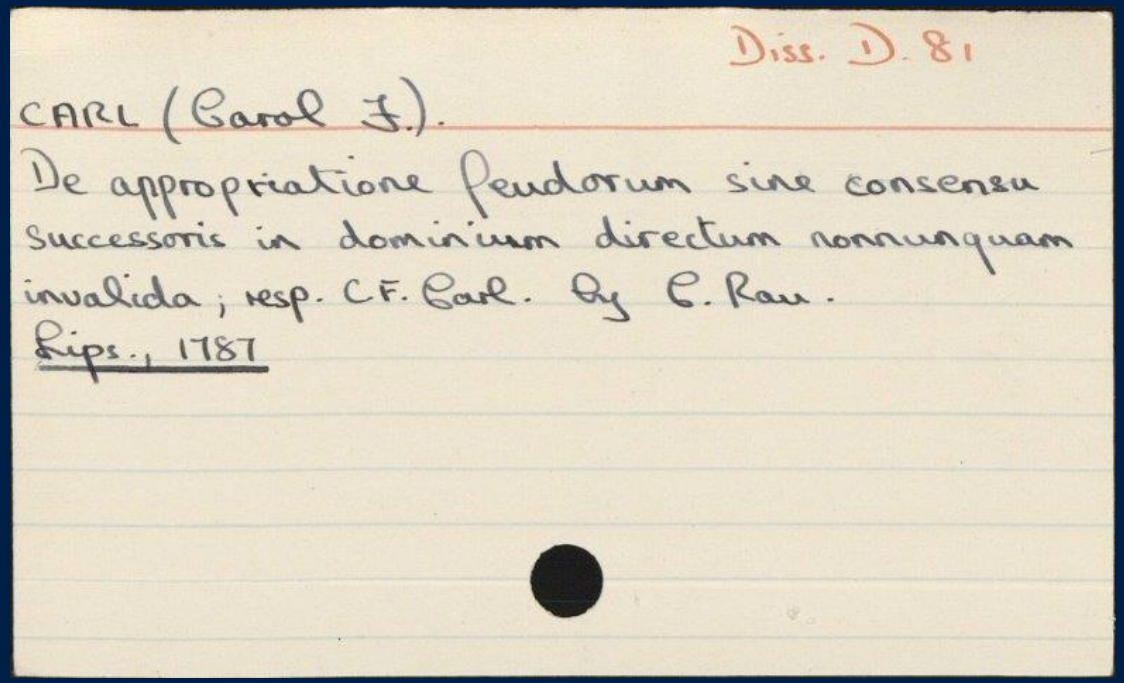
Could AI companies fund our digitisation?

Home 〉 News 〉 Oxford and OpenAI launch collaboration to advance research and education

Oxford and OpenAI launch collaboration to advance research and education

PUBLISHED
4 MAR 2025

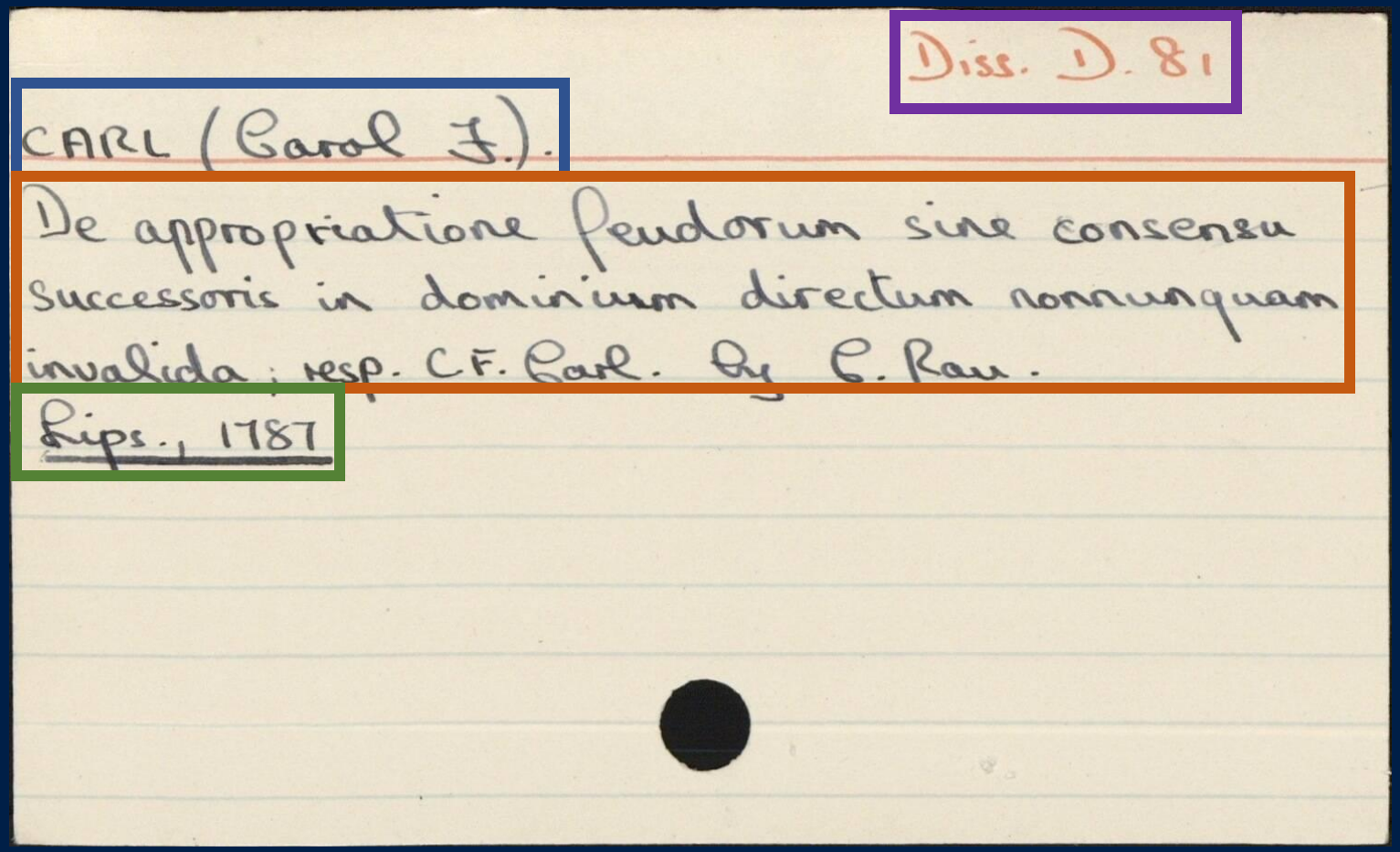
Can we generate metadata with AI?



Does handwriting text recognition work in library workflows? How accurately?

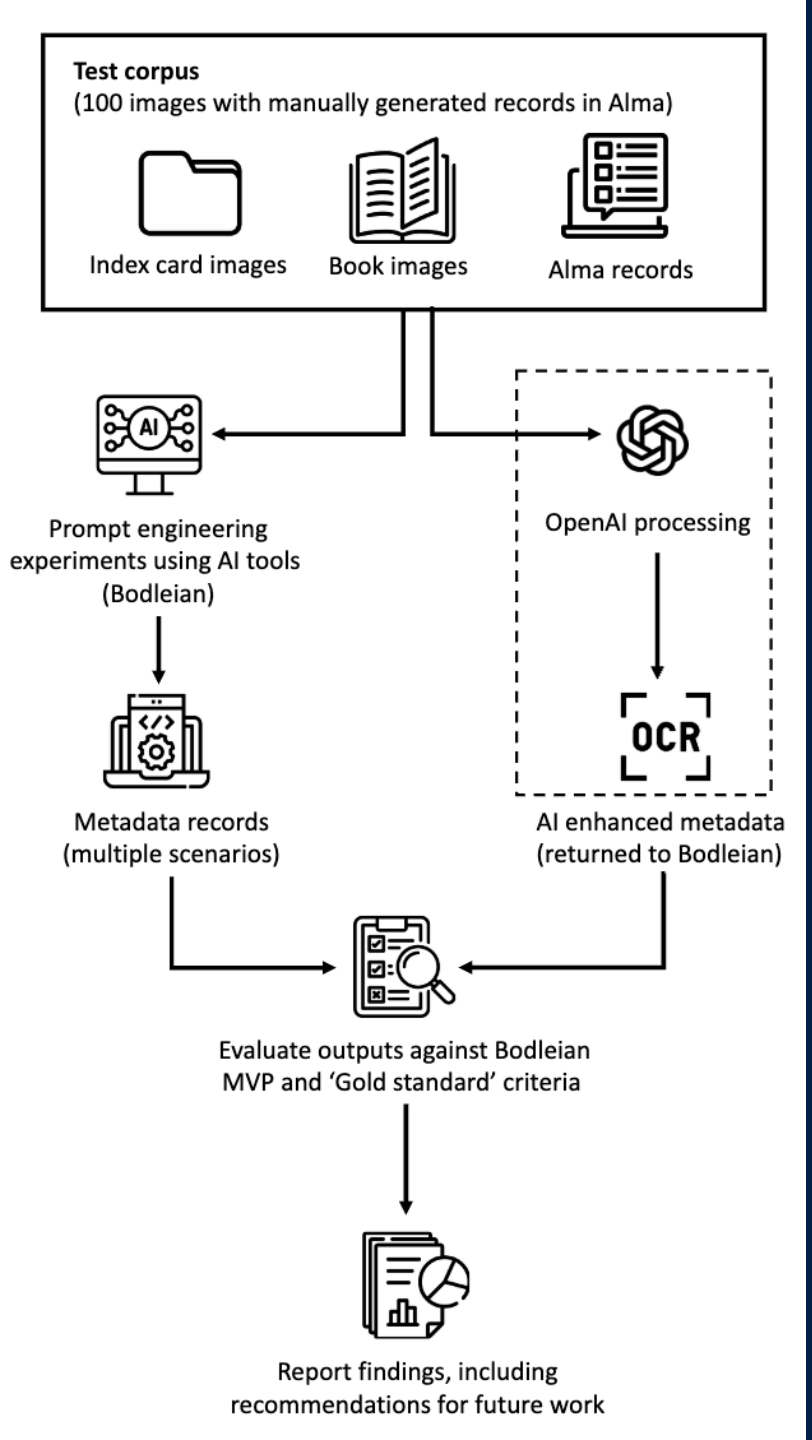
Can AI create structured data (e.g. MARC records) from unstructured data (e.g. full texts)?

Extracting text with AI



"Extract the text information from an image of an index card, ensuring to include only the details specified on the card
Identify the following elements:
Responsible person: Typically starts with a surname in capital letters, followed by a first name in brackets.
Shelfmark: Often found in red ink in the top right corner.
Title: Generally in a European language.
Publication Place: Place
Publication Date: Date
Output Format
Return the information as a json response in a structured format, listing each identified element with its corresponding text from the card.
Notes
Ensure all returned text is verbatim from the image of the index card.
Focus on the specific areas where these elements are typically found for accurate extraction."

What are we hoping to learn?



Bodleian-OpenAI digitisation pilot, Workstream 2 high-level workflow

Challenge 1: Quality and consistency

For example, does accuracy mean:

A. We have accurately transcribed the text on the card

and/or

B. We have accurately identified the metadata fields on the card

and/or

C. The metadata on the original card is correct

Algorithmic archiving project

Will scholars need to access today's algorithmic data for future research?

Challenge 3: How open is too open?

Who owns AI-created works? And with widespread web-scraping, how safe is our own IP?

Can (or should) we turn off web scraping by AI companies?

Top-500 Registered Domains of the Latest Main Crawl

The table below shows the top 500 registered domains (in terms of page captures) of the last main/monthly crawl (CC-MAIN-2025-26). The underlying data is also provided in CSV format, see domains-top-500.csv.

Note that the ranking by page captures only partially corresponds to the importance of domains, as the crawler respects the robots.txt and tries hard not to overload web servers. Highly ranked domains tend to be underrepresented. If you're looking for a list of domain or host names ranked by page rank or harmonic centrality, consider using one of the webgraph datasets instead.

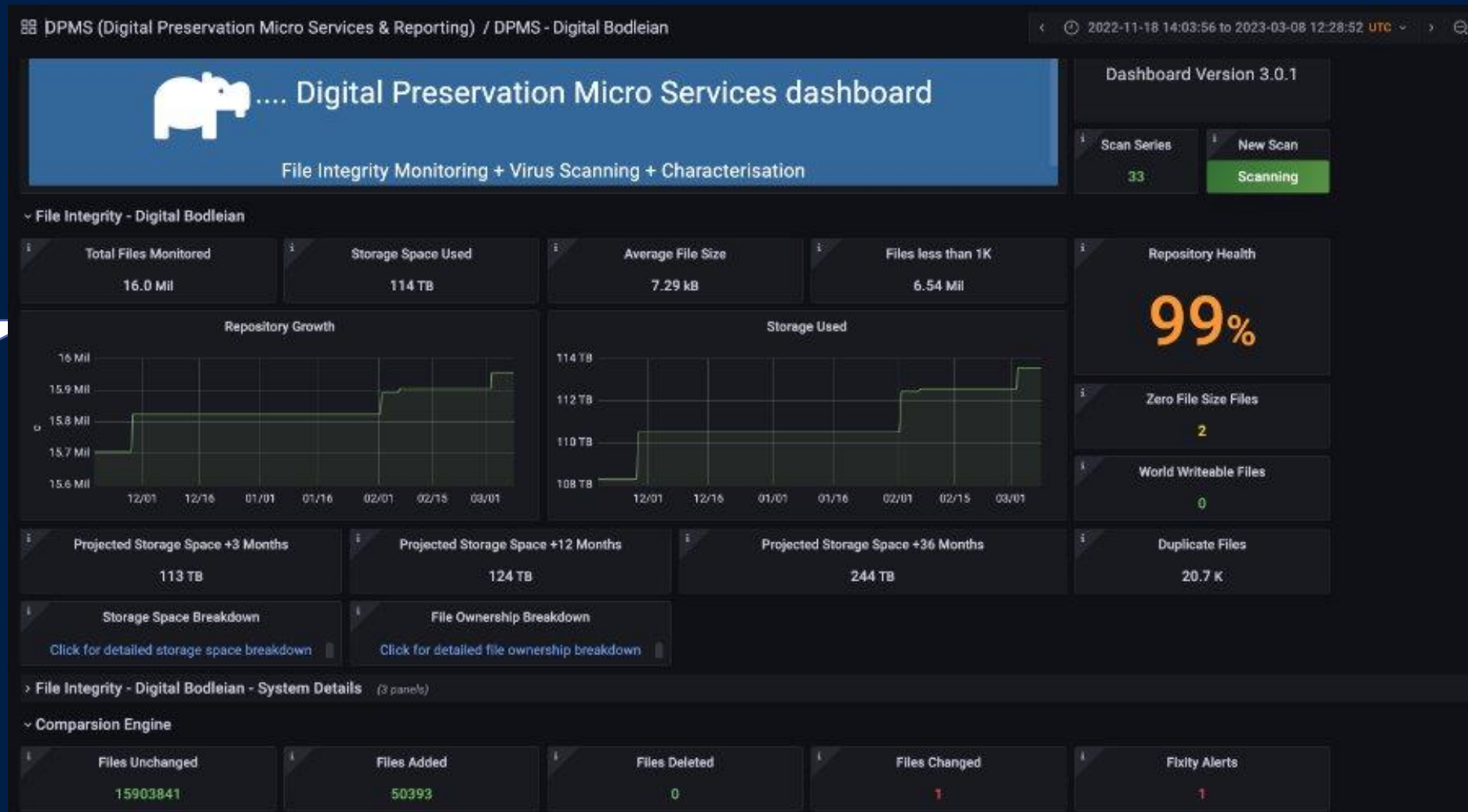
ox

domain	pages	urls	hosts	%pages	%urls
ox.ac.uk	1014316	1012289	1260	0.042518	0.042656

Challenge 4: Keeping up with the pace

What AI skills do students and researchers need?

What AI skills and knowledge do our staff need?



How can we share our learning with each other?

How are you approaching similar questions?

Thank you.

elizabeth.hatfield@bodleian.ox.ac.uk
andrew.bonnie@bodleian.ox.ac.uk