Humanities data

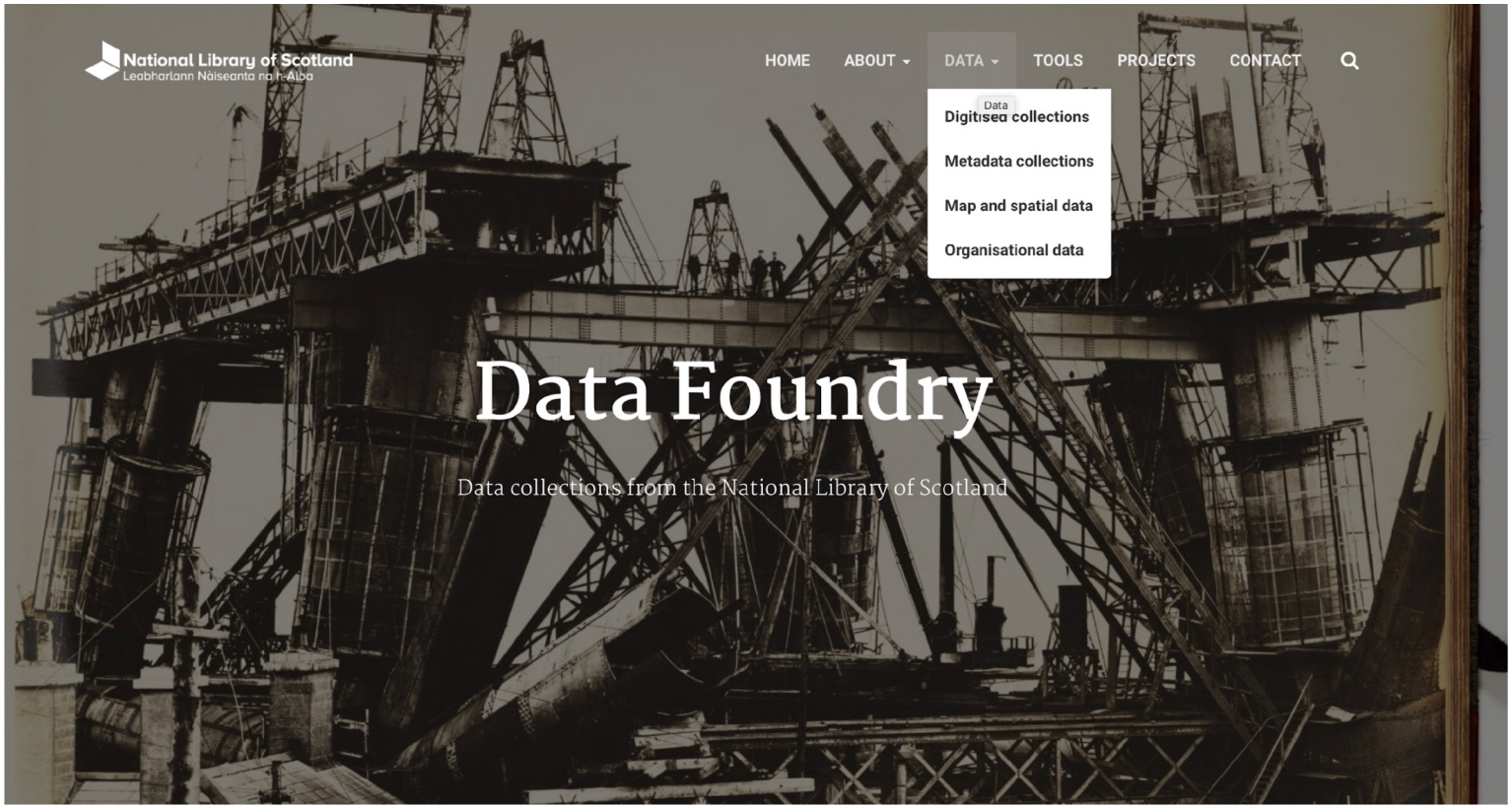
Humanities scholars will sometimes describe elaborate visualizations to me, involving charts and graphs and change over time.

‘Great,’ I respond. ‘Let’s see your data.’

‘Data?’ they say. ‘Oh, I don’t have any data’.

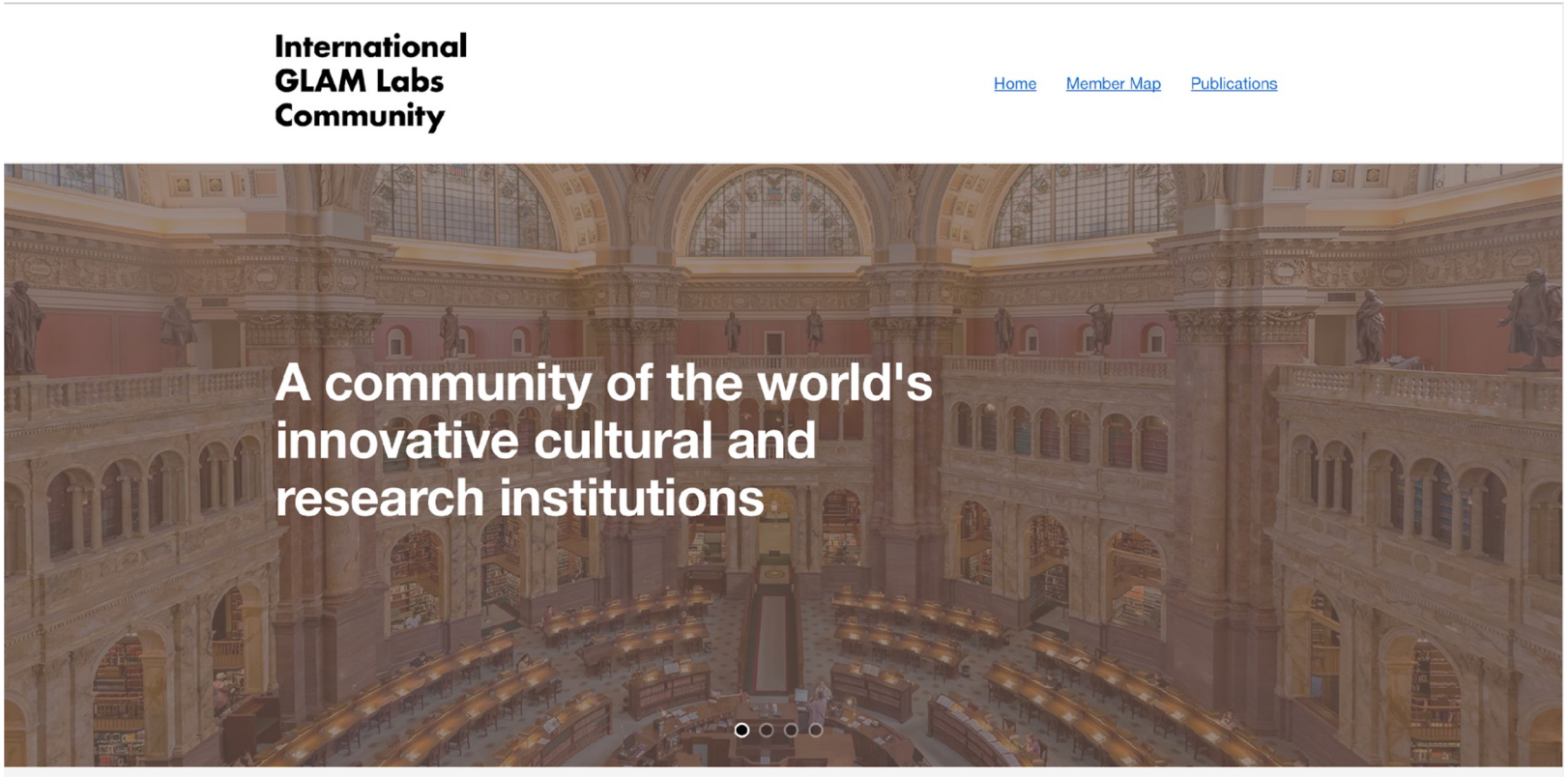
- Miriam Posner, Humanities Data: A Necessary Contradiction. http://miriamposner.com/blog/humanities-data-a-necessary-contradiction/ (15 July 2015)

Curated cultural heritage data



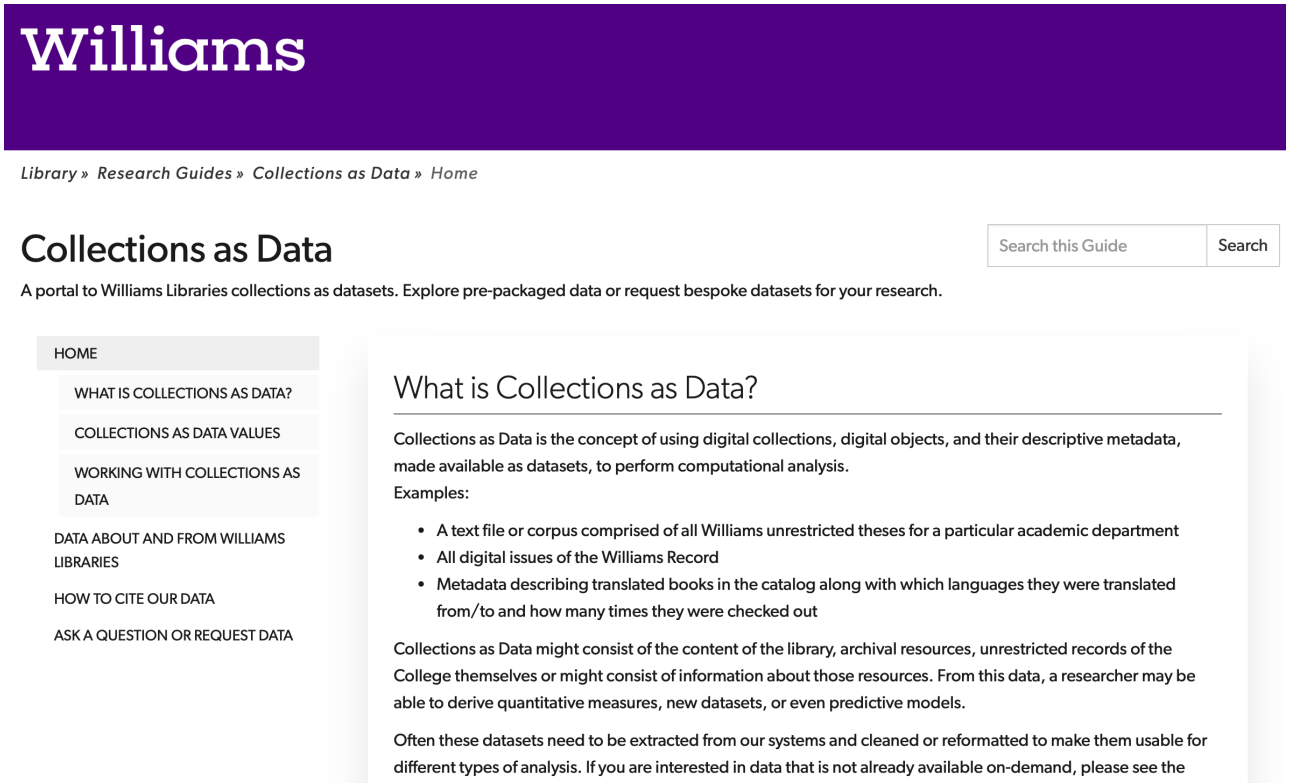
https://data.nls.uk

More curated data



https://glamlabs.io

Collections as data



Catalogues as data

National bibliographies have been identified as a crucial resource for historical research on the publishing landscape, but using them requires addressing challenges of data quality, completeness, and interpretation. We call this approach bibliographic data science.

Lahti, Leo, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. "Bibliographic Data Science and the History of the Book (c. 1500–1800)." *Cataloging & Classification Quarterly*57, no. 1 (2019): 5-23.

Catalogues as data

The use of bibliographic metadata as a research object has, however, proven to be challenging […] [The] Primary motivation for cataloging has been to preserve as much information of the original document and its physical creation as possible. This includes potential errors caused by the printer. If, for instance, a place name is wrongly spelled on the title page, for cataloging purposes it is relevant also to preserve that misspelling. For anyone desiring to work on quantitative approach to bibliographic metadata, this is a crucial point to understand and respect.

.

Lahti, Leo, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. "Bibliographic Data Science and the History of the Book (c. 1500–1800)." *Cataloging & Classification Quarterly*57, no. 1 (2019): 5-23.

Against data as such

… we re-examine the intellectual foundations of digital humanities… first and foremost that we reconceive all data as *capta*..

Capta is ‘taken’ actively while data is assumed to be a given, able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact.

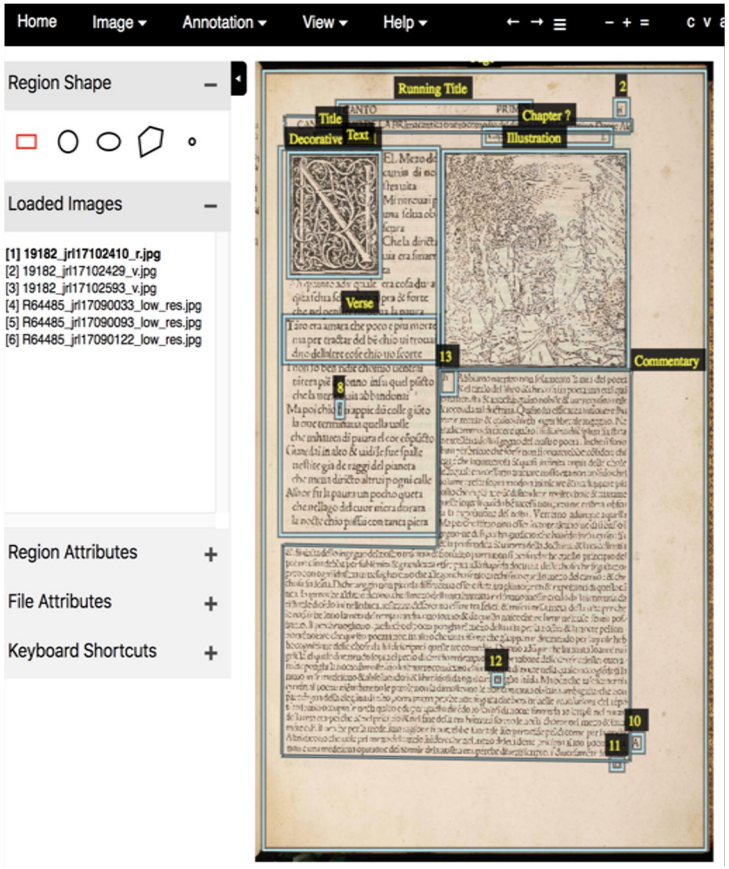
- Johanna Drucker, "Humanities approaches to graphical display." *Digital Humanities Quarterly* 5, no. 1 (2011): 1-21

Texts as data vs. the material text

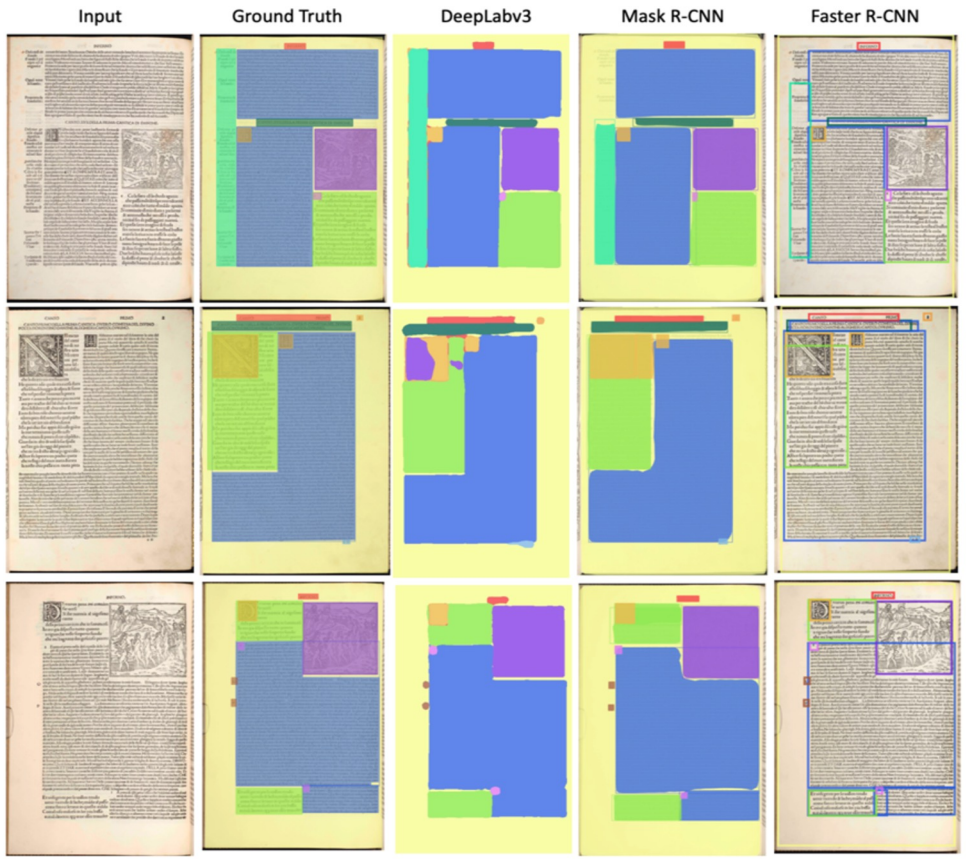
Digital reproductions are central to our work as scholars and teachers [but] many scholars have cautioned against treating digital images as surrogates for the original, citing the loss of texture, size, smell, colour and context. With the proliferation of more and higher-resolution reproductions online, however, subtle but interesting variations between copies at different institutions have become more visible, ironically helping foment copy-specific work in book history and thus spurring on the digitization of even more material.

- Zachary Lesser and Whitney Trettien. "Material / digital." In *Shakespeare / Text: Contemporary Readings in Textual Studies, Editing and Performance*, edited by Claire M. L. Bourne , 402–423. The Arden Shakespeare, 2021

AHRC-funded Envisioning Dante project

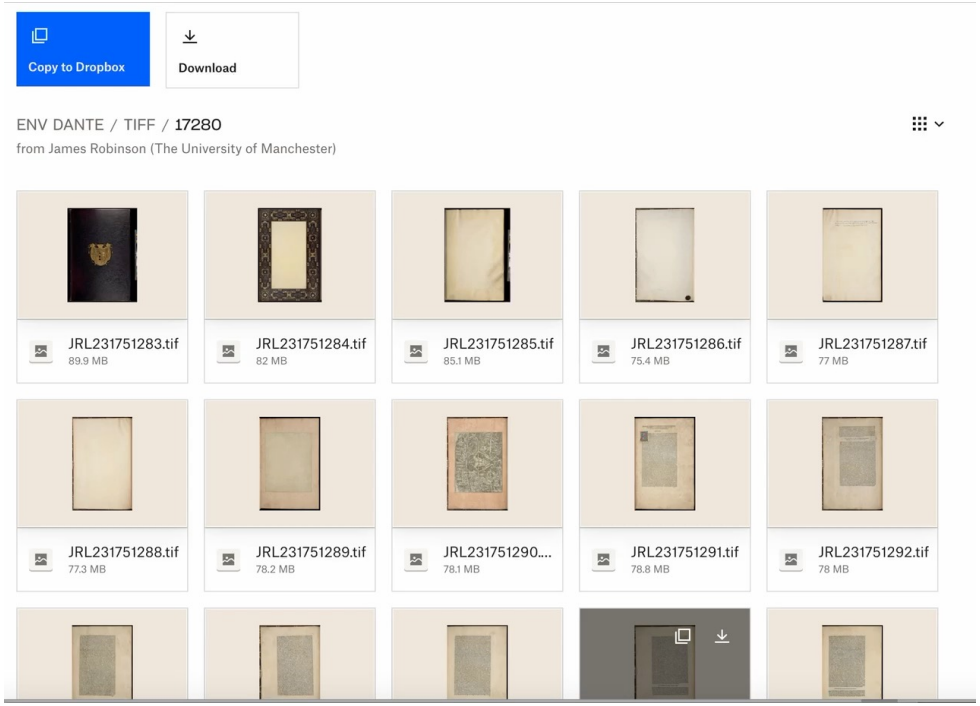


Creating training data with VGG VIA – credit: Charlotte Alton



Testing various machine learning models – Suny Shtedritski, with Andrea Vedaldi

Digitisation and its outputs



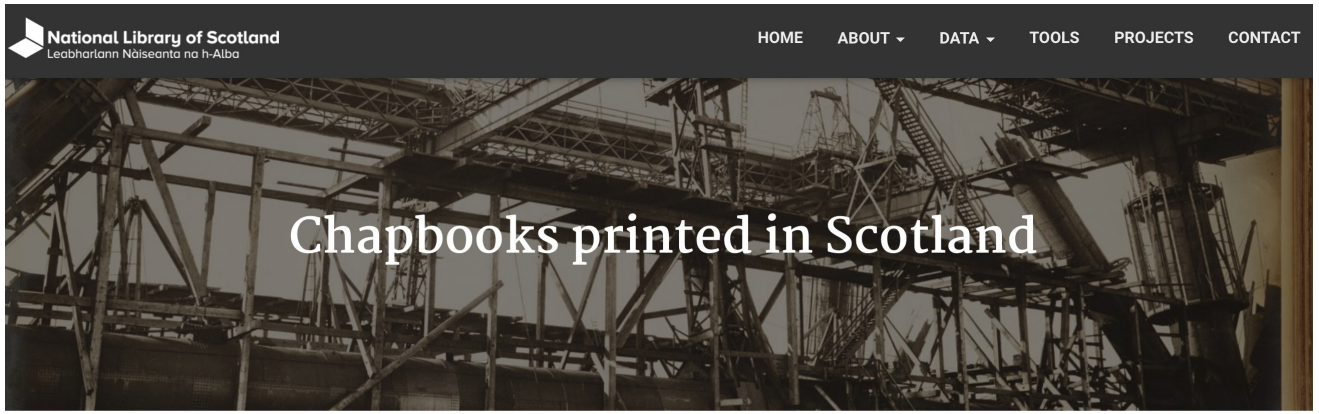
Digitising the page: a caution

Oxford Taylor Institute Library 4 Dante Arch. Fol. It. 1491

We only occasionally see one page of a book at a time; the two pages making an opening are really the unit of the book, and this was thoroughly understood by the old book producers.

- William Morris, 'The Ideal Book', *The Library*, 1893, 179-186 (183)

3000+ digitized chapbooks



Original OCR: no clean-up

47,329 ALTO XML files at page level

47,329 image files

METS metadata files at item level

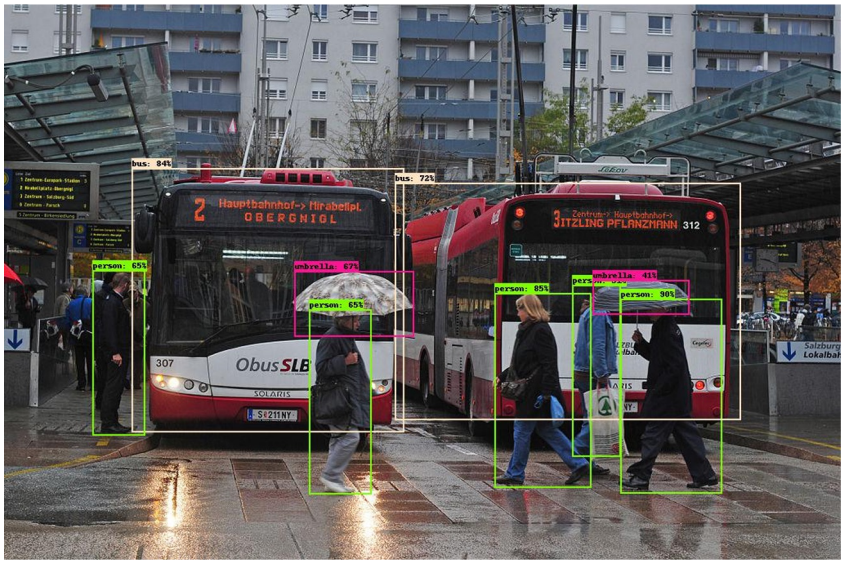
1,432,928 lines and 10,818,763 words

Covers period c1700-c1899

data.nls.uk/data/digitised-collections/chapbooks-printed-in-scotland

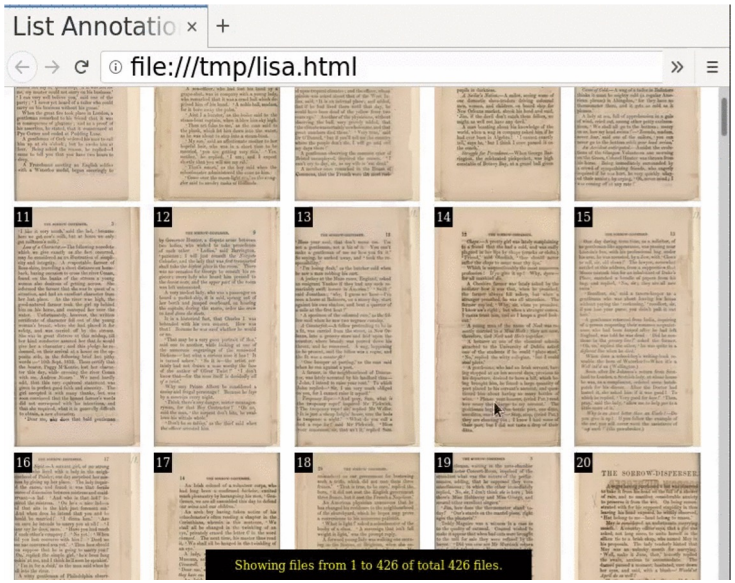
Object Detection

1. We select an object-detection model

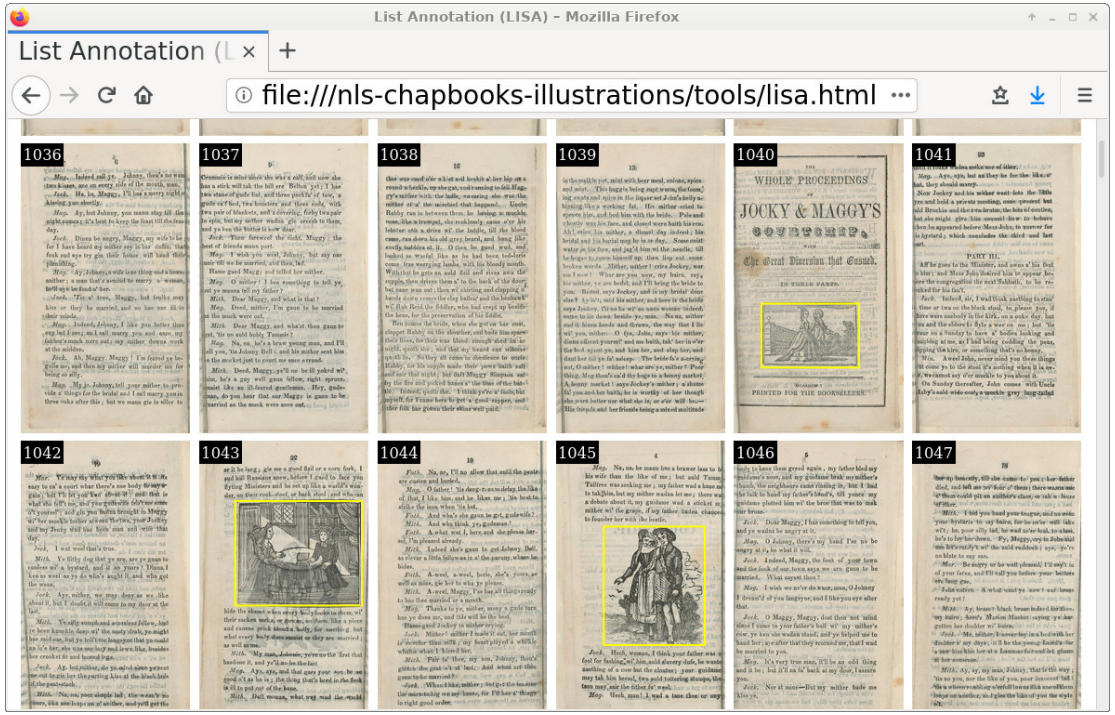


Detecting people, buses and umbrellas with the EfficientDet object detector trained on the COCO dataset

2. We retrain with VGG LISA on book illustrations



3. We successfully predict 3600 illustrations



4. We test and release the retrained model for reuse

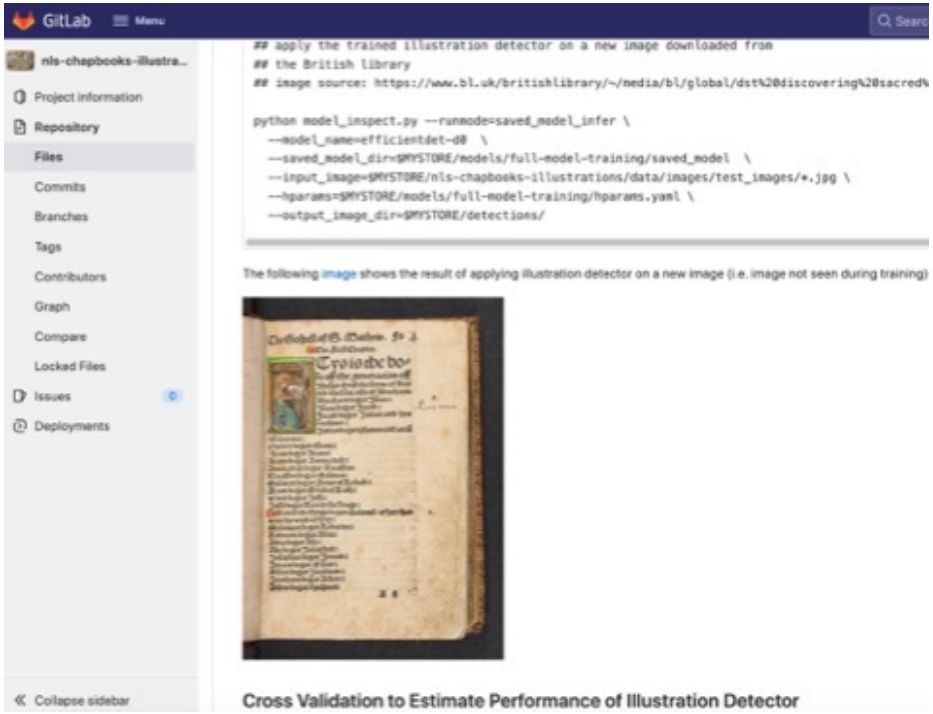
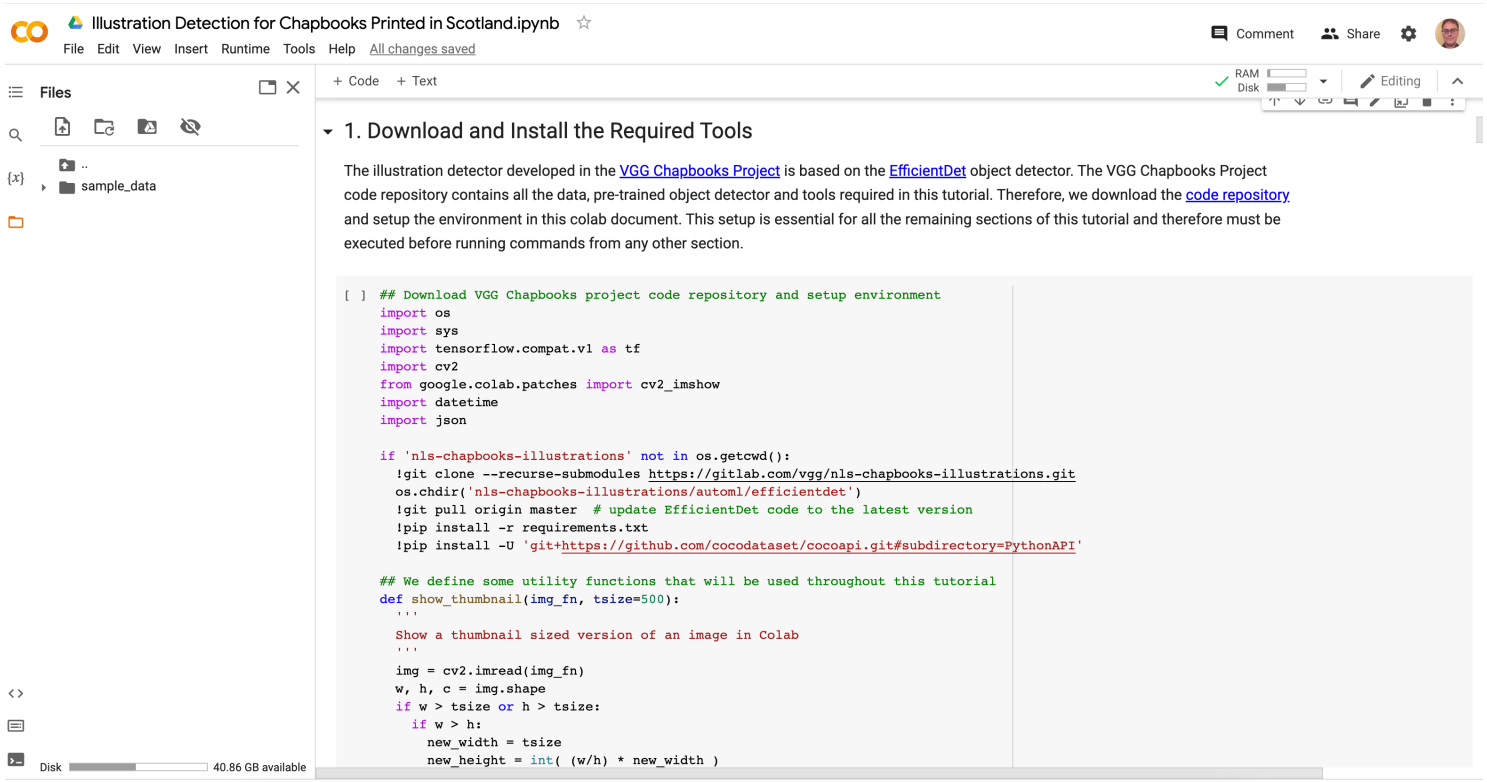


Illustration Detector in a Python notebook



False positives – a printed offset, a provenance stamp and a dog-ear

13002

…iteous cafe he waded through,
and gained the flippery fide,
Where grinning crouds were gather'd round
to mock his fallen pride.
They led him to a neighbouring pump,
to clear his difmal face,
Whence cold and heartleſs home he ſlunk,
involv'd in fore diſgrace.
And many a bill for damage done.
his father had to pay,
Take warning youthful drivers all
from Jehu's firſt eſſay.

FINIS.

6304

We twa hae run about the braes,
When summer days were prime;
But seas between us braid hae roar'd,
Sin' auld langsyne.
For auld lannsyne, &c.

And there's a hand my trusty frien',
And gie's a haud o' thine.
And we'll toom th e stoup to friendship's growth,
And auld langsyne.
For auld langsyne, &c.

And surely ye'll be your pint-stoup,
As sure as 'll be mine,
An' we'll tak a right gude willy-waught,
For auld laggsyne.
For auld langsyne, &c.

FINIS.

13154

Ye wadna miss to get a clout,
I ken he disna fear ye;
Sae haud ye'r tongue and say nae mair,
Set somewere else your fancy:
For as lang's Sandy's to the fore,
Ye never shall get Nancy.

The Sodger Laddie.

My sodger laddie is over the sea,
And he'll bring gold and money to me:
And when he comes hame, he'll make me a lady
My blessing gang wi' my sodger laddie.
My doughty laddie is handsome and brave,
And can as a sodger and lover behave;
True to his country, to love he is steady,
There's few to compare wi' my sodger laddie

Shield him ye ange's, frae death in alarms,
Return him with laurels to my langing arms.
Syne frae all my care ye'll pleasantly free me,
When back to my wishes my sodger ye gie m
O soon may his honours bloom fair on his brow
As quickly they must if he get his due;
For in noble actions his courage is ready,
Which makes me delight in my sodger laddie.

FINIS.

HOME PEOPLE RESEARCH PUBLICATIONS DEMOS DATA SOFTWARE PRACTICALS PROJECTS JOBS INTERNAL

Visual Analysis of Chapbooks Printed in Scotland

Abhishek Dutta, Giles Bergel and Andrew Zisserman

Overview

Chapbooks were short, cheap printed booklets produced in large quantities in Scotland, England, Ireland, North America and much of Europe between roughly the seventeenth and nineteenth centuries. A form of popular literature containing songs, stories, poems, games, riddles, religious writings and other content designed to appeal to a wide readership, they were frequently illustrated, particularly on their title-pages. This paper describes the visual analysis of such chapbook illustrations. We automatically extract all the illustrations contained in the National Library of Scotland Chapbooks Printed in Scotland dataset, and create a visual search engine to search this dataset using full or part-illustrations as queries. We also cluster these illustrations based on their visual content, and provide keyword-based search of the metadata associated with each publication. The visual search, clustering of illustrations based on visual content, and metadata search features enable researchers to forensically analyse the chapbooks dataset and to discover unnoticed relationships between its elements. We release all annotations and software tools described in this paper to enable reproduction of the results presented and to allow extension of the methodology described to datasets of a similar nature.

https://www.robots.ox.ac.uk/~vgg/research/chapbooks/

The problem of sustaining DH data

Project name

Arts and Humanities Data Service: Enabling Digital Resources for the Art and Humanities

Project principal investigator(s)

Sheila Anderson

Decommission Date

September 2017

Archive URL(s)

http://www.ahds.ac.uk/

Additional links

Internet Archive

Overview

The Arts and Humanities Data Service was a UK national service funded by the JISC and AHRC to collect, preserve and promote the electronic resources which result from research and teaching in the arts and humanities.

https://ahds.ac.uk

The problem of sustaining humanities data

UNIVERSITY OF OXFORD

Holding page: British Book Trade Index

The Bodleian Libraries have taken The British Book Trade index resource offline, owing to a change in guidance within the University relating to cyber security.

The British Book Trade Index resource has therefore been taken offline as a precaution, in the light of the new guidance, while we develop new approaches to being able to support and deliver it.

We acknowledge how disruptive this has been to the many scholars and communities who use this resource. Alternative ways to access The British Book Trade Index resource are below, while we determine routes and funding to take the resource forward.

More on the background here: Bodleian Service Updates

An export of the BBTI database and John Feather's Bibliography is available via the Oxford University Research Archive.

The above dataset has been redeveloped as a stand-alone site by Prof. Janelle Jansted and Martin Holmes at University of Victoria.

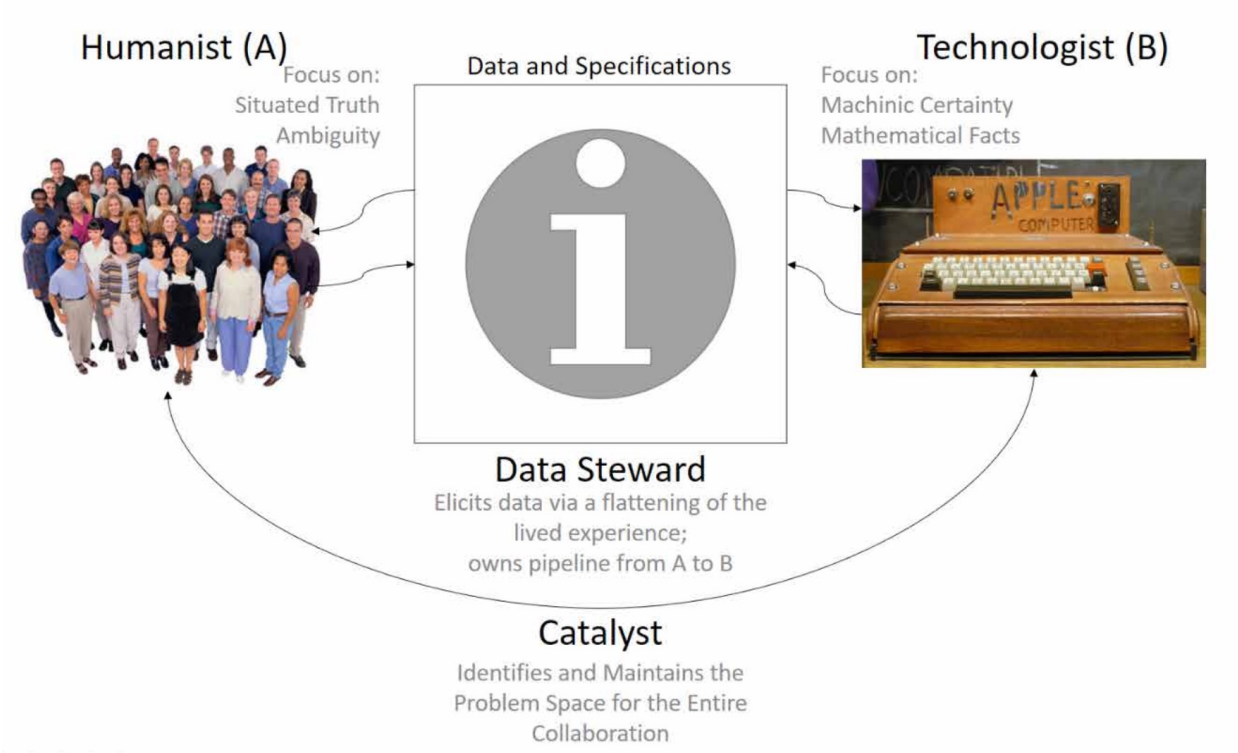
The dataset has also been redeveloped as a stand-alone site by Prof. Allison Muri of the University of Saskatchewan.

Users may also wish to consult the Dicey & Marshall catalogue.

Please contact digitalsupport@bodleian.ox.ac.uk with any questions.

https://bbti.bodleian.ox.ac.uk

Data stewardship



We propose that effective collaborations in Digital Art History require more than just a humanist and a technologist to succeed. Indeed, we find that there are four different roles that need to be filled: Humanist, Technologist, Data Steward, and Catalyst.

It is important to note here that we are talking about roles, not people. Two or more of these roles can be performed by the same person.

Berg-Fulton, Tracey, Alison Langmead, Thomas Lombardi, David Newbury, and Christopher Nygren. "A Role-Based Model for Successful Collaboration in Digital Art History." *International Journal for Digital Art History: Issue 3, 2018: Digital Space and Architecture* 3 (2019): 153

The role of the data steward

The Data Steward is responsible for ensuring that the essential character of the historical evidence is not lost throughout the process of converting primary source material into computable information, and also that this data will be suitable for the technological processes that it will undergo.

By designating this role as the party responsible for surfacing both technological and historical assumptions throughout the process of preparing the data for computation, we make explicit the requirement to observe and discuss the constraints and limitations of this entire process.

Berg-Fulton, Tracey, Alison Langmead, Thomas Lombardi, David Newbury, and Christopher Nygren. "A Role-Based Model for Successful Collaboration in Digital Art History." *International Journal for Digital Art History: Issue 3, 2018: Digital Space and Architecture* 3 (2019): 153

Further reading

Alan Galey, ‘The Enkindling Reciter: E-Books in the Bibliographical Imagination,’ *Book History* 15 (2012): 210-47

Martin Paul Eve, ‘You have to keep track of your changes’: The Version Variants and Publishing History of David Mitchell's Cloud Atlas, *Open Library of Humanities* 2, no. 2 (2016): 1-34

Bonnie Mak, ‘Archaeology of a Digitization,’ *Journal of the Association for Information Science and Technology* 65, no. 8 (2014): 1515–26

Zachary Lesser and Whitney Trettien, ‘Material / digital,’ In *Shakespeare / Text: Contemporary Readings in Textual Studies, Editing and Performance*, edited by Claire M. L. Bourne, (London: The Arden Shakespeare, 2021), 402–423

Whitney Trettien, ‘The History of the Early Modern Book in the Digital Age,’ in Adam Smyth (ed.), *The Oxford Handbook of the History of the Book in Early Modern England*, Oxford Handbooks (2023; online edn, Oxford Academic, 18 Sept. 2023)

Ryan Cordell, ‘”Q i-tjb the Raven”: Taking Dirty OCR Seriously,’ *Book History* 20 (2017): 188–225

Matthew Kirschenbaum and Sarah Werner. "Digital scholarship and digital studies: The state of the discipline." *Book History* 17, no. 1 (2014): 406-458