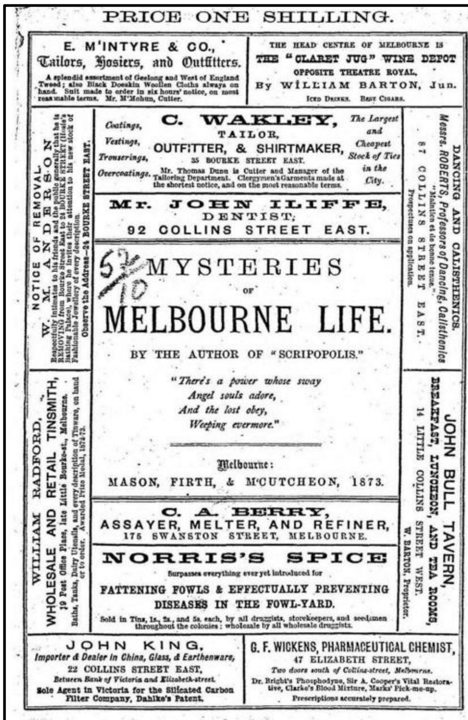
Thank you for the invitation to speak.

I will approach the question by describing the mechanisms organisations have used to open and connect data, then I will look at some of the positive outcomes that resulted from their actions. This is not a technical talk about different acronyms, it's about connecting people to our shared heritage.

Overview

- Definitions
- How - examples of linked data
- How - examples of open data
- Why - what happens when you connect and open data?

Definitions



PRICE ONE SHILLING.

E. M'INTYRE & CO.,
Tailors, Hosiers, and Outfitters.
A splendid assortment of Geelong and West of England Tweed; also Black Doeskin Woollen Cloths always on hand. Suit made to order in six hours' notice, on most reasonable terms. Mr. M'Mohun, Cutter.

THE HEAD CENTRE OF MELBOURNE IS
THE "CLARET JUG" WINE DEPOT
OPPOSITE THEATRE ROYAL,
By WILLIAM BARTON, Jun.
Iced Drinks. Best Cigars.

NOTICE OF REMOVAL.
W. M. ANDERSON
Respectfully intimates to his friends and the public generally that he is REMOVING from Bourke Street East to 24 BOURKE STREET (Hosie's Bathing Palace), where he invites their attention to his new stock of Fashionable Jewellery of every description.
Observe the Address—24 BOURKE STREET EAST.

Coatings, Vestings, Trouserings, Overcoatings.
C. WAKLEY,
TAILOR,
OUTFITTER, & SHIRTMAKER,
35 BOURKE STREET EAST.
Mr. Thomas Dunn is Cutter and Manager of the Tailoring Department. Clergymen's Garments made at the shortest notice, and on the most reasonable terms.
The Largest and Cheapest Stock of Ties in the City.

DANCING AND CALISTHENICS.
Messrs. ROBERTS, Professors of Dancing, Calisthenics
"Maintien et de bonne tenue."
87 COLLINS STREET EAST.
Prospectuses on application.

Mr. JOHN ILIFFE,
DENTIST,
92 COLLINS STREET EAST.

MYSTERIES
OF
MELBOURNE LIFE.

BY THE AUTHOR OF "SCRIPOPOLIS."

"There's a power whose sway
Angel souls adore,
And the lost obey,
Weeping evermore."

Melbourne:
MASON, FIRTH, & M'CUTCHEON, 1873.

WILLIAM RADFORD,
WHOLESALE AND RETAIL TINSMITH,
19 Post Office Place, late Little Bourke-st., Melbourne.
Baths, Tanks, Dairy Utensils, and every description of Tinware, on hand or to order. Awarded Prize Medal, 1872-73.

JOHN BULL TAVERN,
BREAKFAST, LUNCHEON, AND TEA ROOMS,
14 LITTLE COLLINS STREET WEST.
W. BARTON, Proprietor.

C. A. BERRY,
ASSAYER, MELTER, AND REFINER,
175 SWANSTON STREET, MELBOURNE.

NORRIS'S SPICE
Surpasses everything ever yet introduced for
FATTENING FOWLS & EFFECTUALLY PREVENTING DISEASES IN THE FOWL-YARD.
Sold in Tins, 1s., 2s., and 5s. each, by all druggists, storekeepers, and seedsmen throughout the colonies; wholesale by all wholesale druggists.

JOHN KING,
Importer & Dealer in China, Glass, & Earthenware,
22 COLLINS STREET EAST,
Between Bank of Victoria and Elizabeth-street.
Sole Agent in Victoria for the Silicated Carbon Filter Company, Dahlke's Patent.

G. F. WICKENS, PHARMACEUTICAL CHEMIST,
47 ELIZABETH STREET,
Two doors south of Collins-street, Melbourne.
Dr. Bright's Phosphodyne, Sir A. Cooper's Vital Restorative, Clarke's Blood Mixture, Marks' Pick-me-up.
Prescriptions accurately prepared.

A splendid assortment of Gceloag and West of England. Tweed ; also Black Doeakin Woollen Cloths alwaya on hand. Snit made to order in six hoars' notice, on most reaainable terms. Mr. M'Mohon, Cutter.

Mysteries of Melbourne life
by Cameron, Donald, 1848?-1888.

Published 1873
Usage Public Domain Mark 1.0
Topics Australia -- Fiction

What kinds of data are we talking about? It might just be metadata - the title, author, dates, places… Or 'data' might include digital images of pages or objects. Or it might include full text transcriptions, with or without the optical character recognition errors shown here. Any of these can be offered as collections of metadata, of text, of images, for reading individually or mining as a dataset.

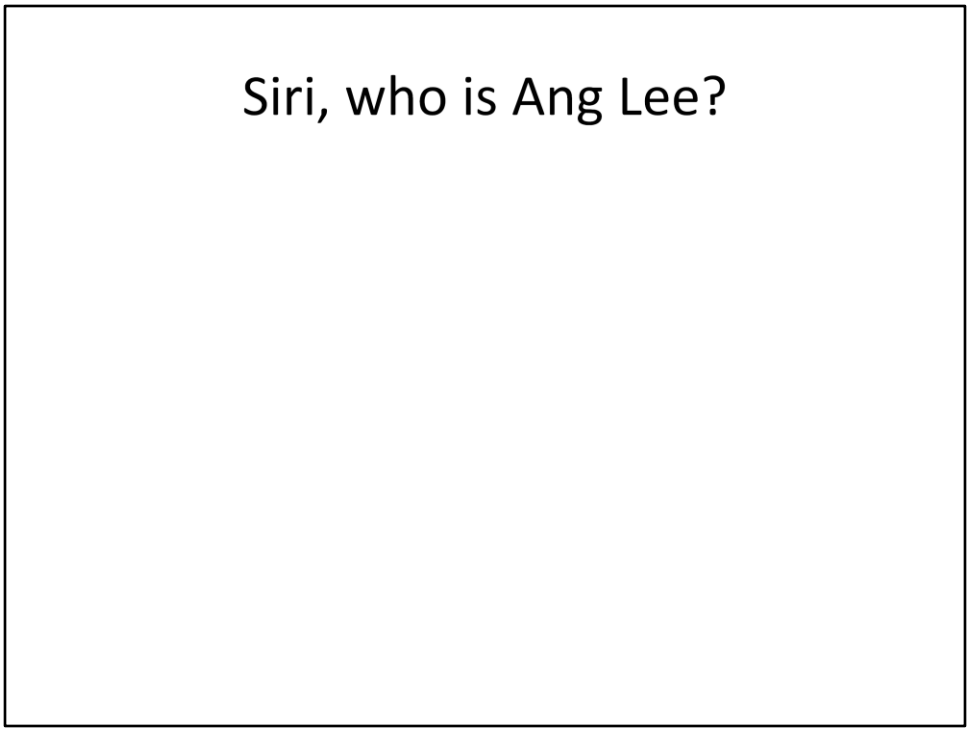
Image, data. https://archive.org/details/MysteriesOfMelbourneLife

Open and connected data

Open data: machine-readable data made available under an open licence. Ideally published in a non-proprietary format (CSV, XML, JSON, RDF, SPARQL). Enables commercial and creative re-uses.

Connected data: uses links (URLs) to describe concepts. Ideally published at a stable address so others can use it as a reference. Enables interoperability and collaboration.

I could go into more technical detail, but this is a high-level overview. Not about acronyms. Aim to add enough structure to let people know what your collections are about.



Ang Lee might be a common name, but you probably know who I mean. But would a computer?

Siri, who is Ang Lee?

Image: nicolas genin

This guy!

Siri, who is Ang Lee?

Image: nicolas genin

https://www.wikidata.org/wiki/Q160726

occupation	film director 0 references screenwriter 0 references writer 0 references

Getty ULAN (Union List of Artist Names)
ID: 500334116
Identifying URL:
http://vocab.getty.edu/ulan/500334116

But while people think in 'things', computers can't find meaning in strings of text unless we give them more useful information. So when we identify Ang Lee with a link, the computer can 'understand' that the string of letters is describing a person

Siri, who is Ang Lee?

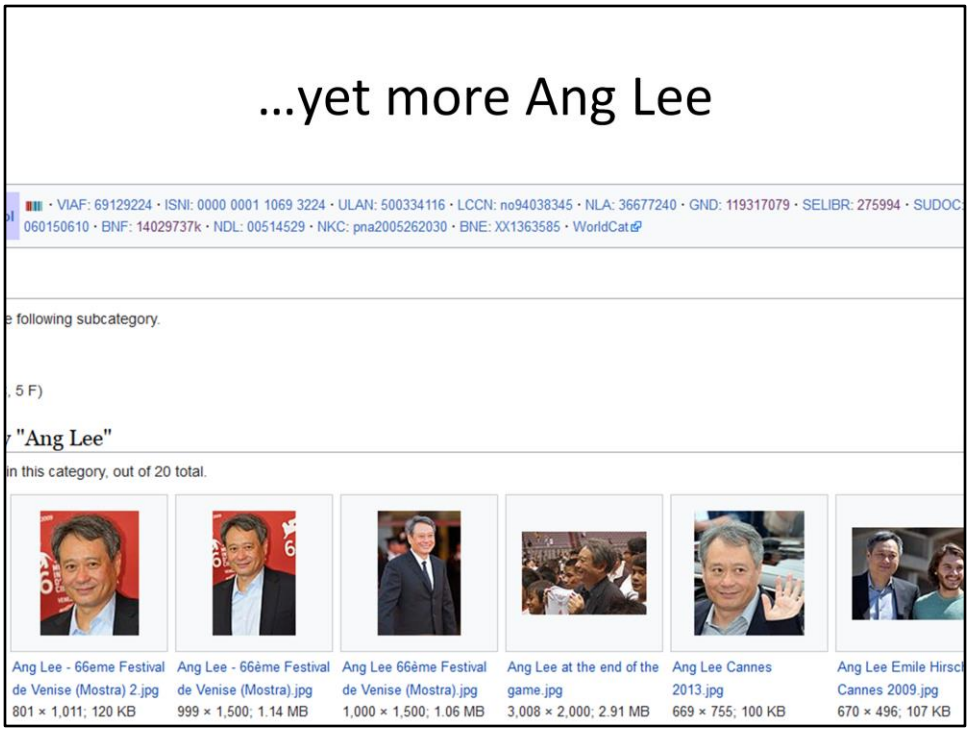
Image: nicolas genin

Titles:
All the better for her (1 min 17 s)
All the delights of the season (1 min 14 s)
Banquete de boda, El
Brokeback mountain
Burza lodowa
Cabalga con el diablo
Chevauchée avec le diable
Combe magna (2 min 59 s)
Comer beber amar
Crouching tiger, hidden dragon
Devonshire (1 min 04 s)
dream (2 min 30 s)
Eat, drink, man, woman
En terreno vedado
Excellent notion (1 min 39 s)
Felicity (1 min 22 s)
Fornuft og følelse
Garçon d'honneur
Grant me an interview (1 min 05 s)
Hôtel Woodstock
Hsi yen, 1993:
Hsi Yen (Pel·lícula cinematogràfica).
Hulk
Ice storm., The
Lust, caution

http://www.isni.org/isni/0000000110693224

Once you have one identifier, you can confidently map to others, and take in their data… This ISNI record includes a comprehensive list of Lee's films because the site understands the link between a creator and their works.

…yet more Ang Lee



Wikimedia Commons includes lots of common identifiers, so you can find more material about Lee.

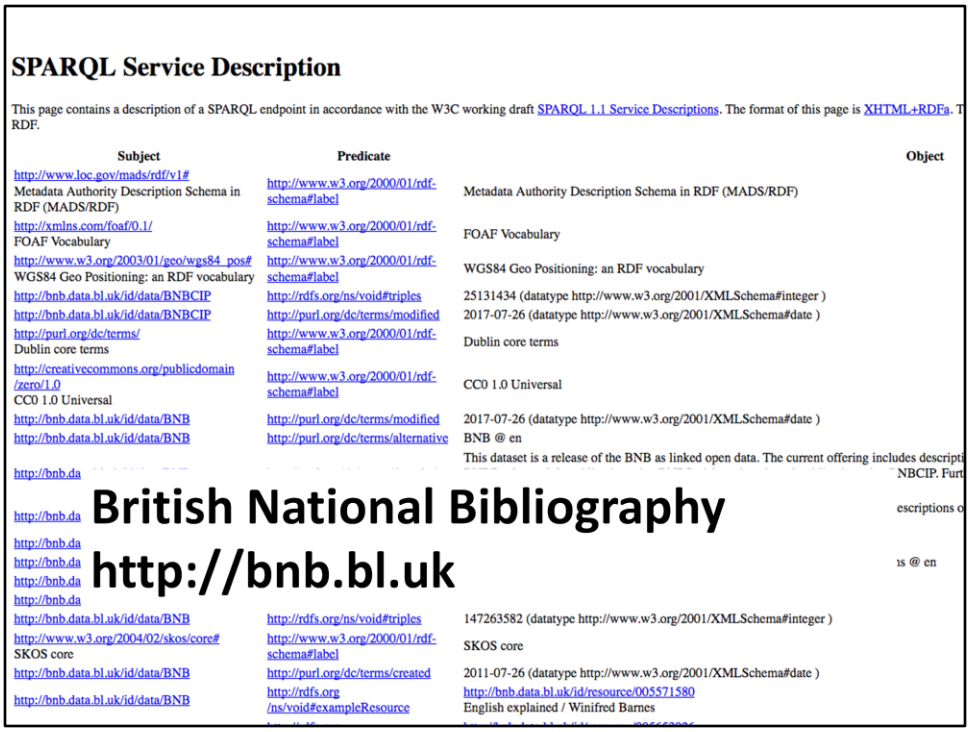
From 'strings' to 'things'

So now we can say:

Ang Lee (https://www.wikidata.org/wiki/Q160726) is a film director (https://www.wikidata.org/wiki/Q2526255)

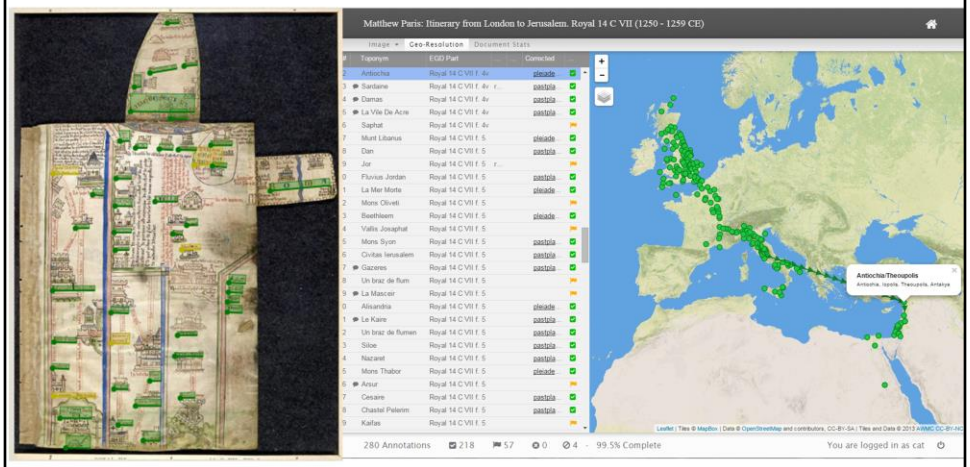
…and it'll mean something to humans and to computers

How - examples of linked data



Books and journal titles published in the UK since the 1950s. Used by libraries to get new records. Available in a range of linked data formats including this query service which describes material in 'triples'. Triples are a format for making statements like 'This book's title is 'Welcome to Taiwan' with semantic information included. Triples allow more complex queries without complicated relational database structures.

'linked data without the pointy brackets'



Pelagios: Enabling Linked Ancient Geodata http://www.pelagios.org

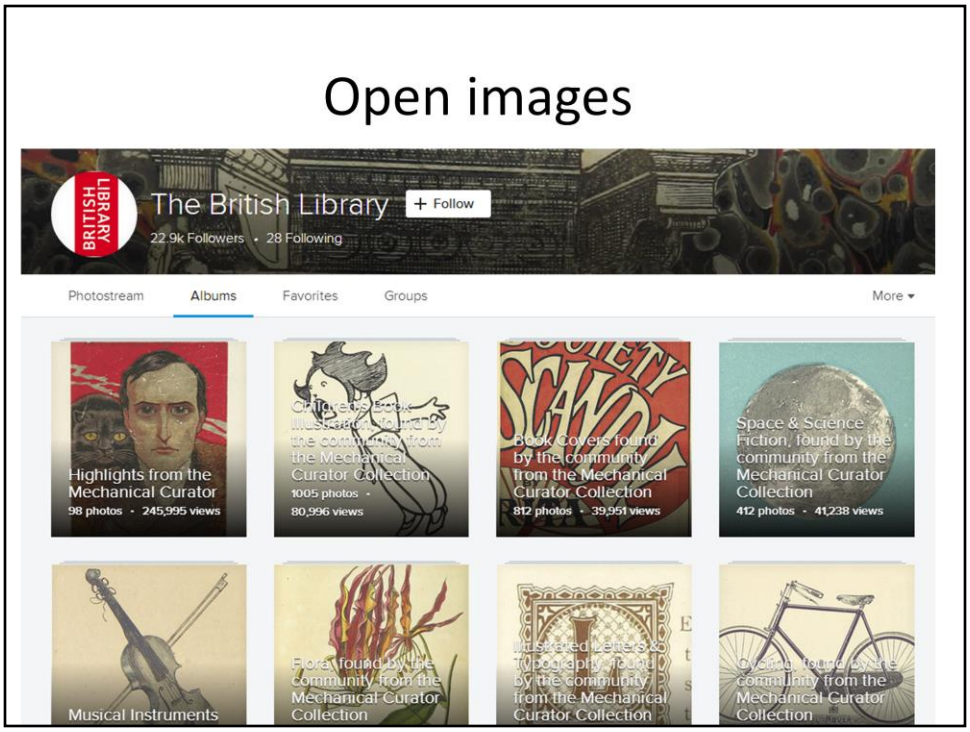
Lots of people find linked open data technologies intimidating. Projects like Pelagios allow placenames in documents and maps to be annotated with linked open data identifiers - you just draw around the word and match it to a list of potential placenames. These identifiers mean the items are more easily linked to other collection items that mention the same places.

The logos are institutions that have shared images freely on Flickr Commons

CSV

- Plain-text, comma-separated format files
- Spreadsheet-style format is familiar to many
- Low barrier to entry - easy to play with

- Can't be updated once downloaded
- Hard to accept changes e.g. if people clean data or add links to connect with other datasets



Since 2008, it's been possible to share freely licenced images on Flickr Commons. The BL put one million images extracted from 19th C books on Flickr Commons. This story is periodically picked up by the media…

Downloadable datasets

Digitised Books - Images identified as Embellishments. c. 1510 - c. 1900. JPG.

Image, pg 145: 'Ballads and Songs of Lancashire chiefly older than the 19th Century. Collected, compiled, and edited, with notes, by J. H., etc. L.P' by John Antiquary Harland, 1865

The dataset comprises c. 41,6951 images identified as 'Embellishments' from the British Library's Flickr Commons collections, dating between c. 1510 - c. 1900. The images were algorithmically gathered from 49,455 digitised books, equating to 65,227 volumes (25+ million pages), published between c. 1510 - c. 1900. The books cover a wide range of subject areas including philosophy, history, poetry and literature. The images are in .JPEG format.

Content Notes:

Full title:	Digitised Books - Images identified as Embellishments. c. 1510 - c. 1900. JPG.
Dataset creator:	British Library
Dataset publisher(s):	British Library
Dataset date:	2014
Range of years covered:	c. 1510 - c. 1900
DOI:	https://doi.org/10.21250/db17

Download:

dig19cbooks-embellishments.zip (link provided by dataset creator)

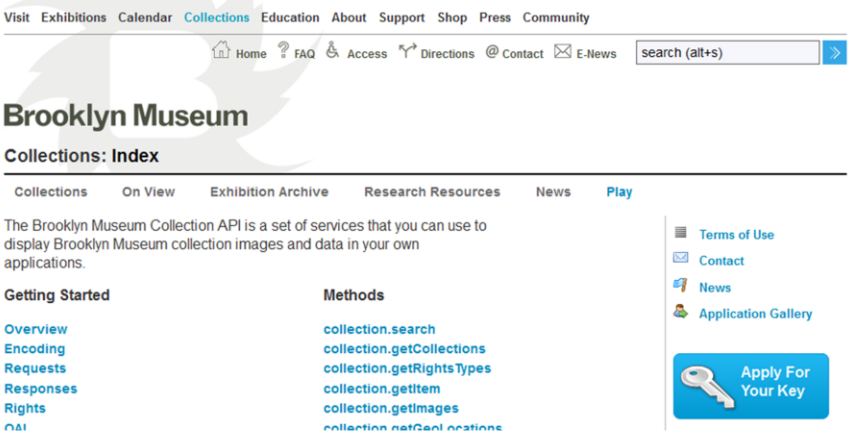
43.4GB .ZIP file, comprising of 416,951 .jpeg images in 316 folders

CC0 1.0 Public Domain

If you have any queries about this dataset or about this website, please email digitalresearch@bl.uk.

Collecting together datasets under data.bl.uk.

APIs



The first museum API was released by the Brooklyn Museum of art in 2009. APIs provide a way to request specific information from a dataset and use the results in another programme. This means other people can make programmes based on your data.

How to open your data?

The basics: find digital or digitisable data that can be openly licenced and doesn't include personal data. Choose the simplest, most sustainable and most usable format.

Build for success: choose data people want to use. Document it well. Test its usability. Publicise it and respond to feedback. Celebrate and share interesting uses.

To help you think about which method is right for you… Not a technical bit.

Narrow down from all the information you have to what's shareable. E.g. copyright and data protection ok, digital or able to be digitised.

Work out what data you have in a shareable state – is it metadata, digitised images, full text? Work out who might be interested in using it – visual designers, historians, computational linguists, literature scholars? What formats are in the sweet spot of being easy to support (for you) and easy to use (for them)? Think about whether you'll want to update it often or only once. Can you publish directly from systems in daily use to ensure that the data is maintained?

Making sure your data is used

Competitions and events

BL Labs Awards rewards Research, Commercial, Artistic, Teaching/Learning work with digital collections

Off the Map encourages game design students with exhibition themes

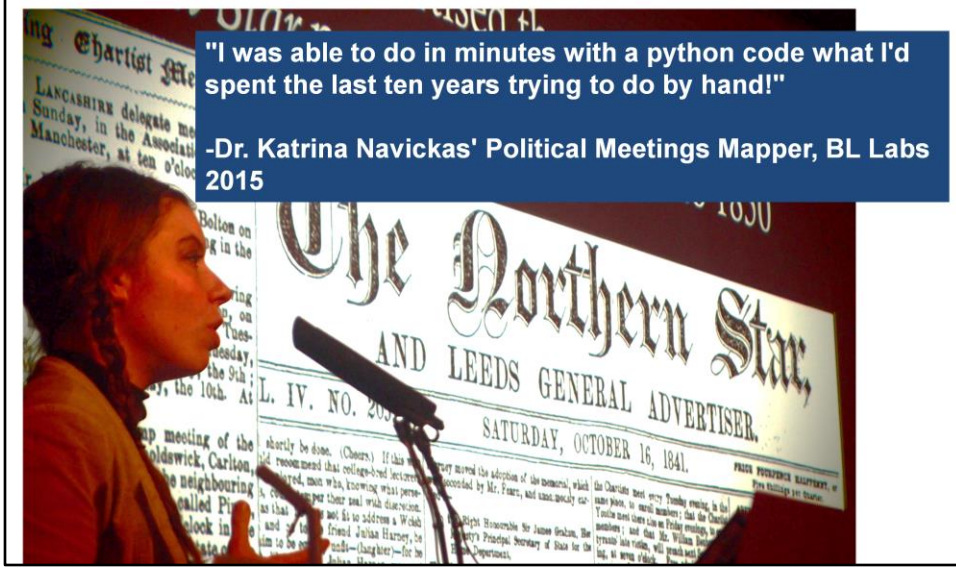
Hackathons: great for feedback

http://gamecity.org/alices-adventures-off-the-map-winners-announced/

Competitions seem effective way to motivate participation - time-limited, clear expectations re what to submit, clear rewards and recognition. Off the Map, run by colleague Stella Wisdom, explores how British Library digital collections can be used in creative ways. Opportunity for game, design students in the UK to showcase their talents to industry; hopefully lead to engagement with new audiences.

BL Labs Awards recognise work done with digital collections (in research, commercial, artistic, teaching/learning categories); Competition winners got to work closely with Labs team for several months to make their DS project idea a reality.

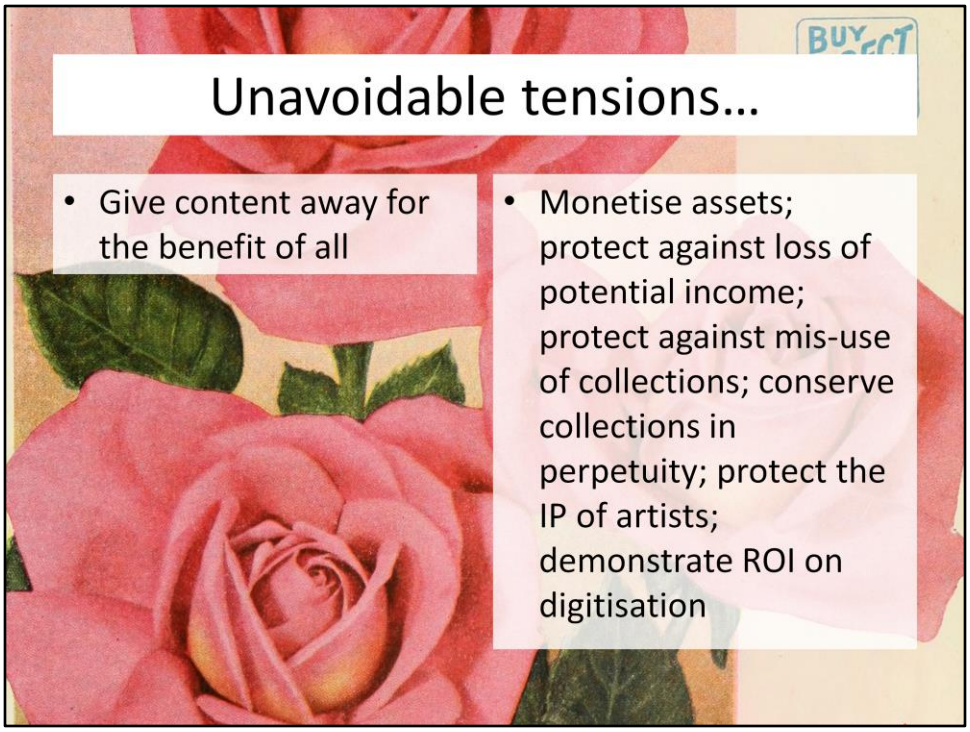
Create and share case studies



Great thing about competitions is they help create case studies. Dr Katrina Navickas was a Labs Competition winner whose story particularly resonates with 'traditional' historians and academics.

Katrina used text mining techniques to map weekly political meetings as advertised in the Northern Star newspaper between 1841 and 1845. Previously she manually created a dataset by taking notes in archives.
http://politicalmeetingsmapper.co.uk/maps/methods

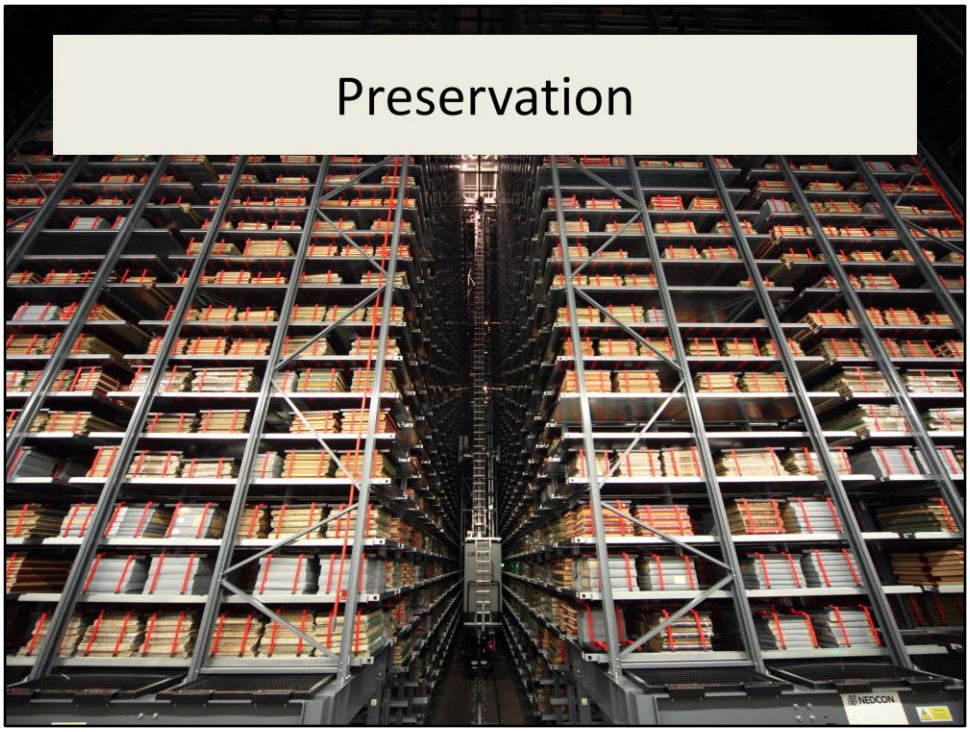
Labs team, usually with a digital curator, run workshops and activities at universities around the country, to encourage re-use of digital collections. Find a format suitable for your organisation - the point is to actively promote the data to people



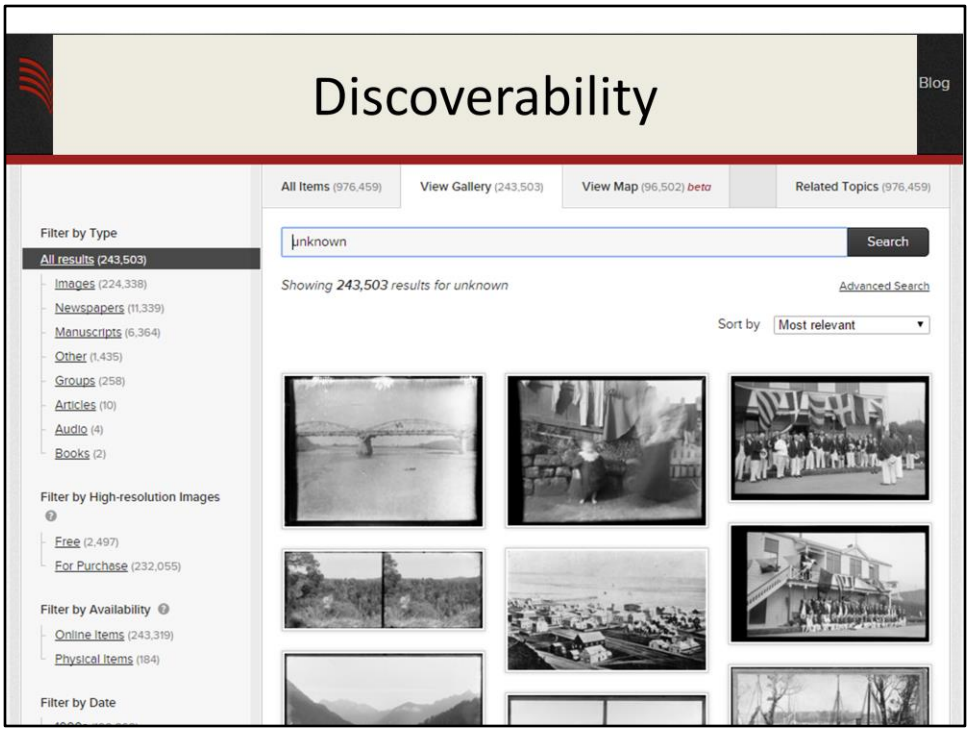
It's not all a bed of roses… Here are some contradictory things cultural heritage institutions (in the UK) are told they must do. How do you balance these requirements?

IP - intellectual property; ROI - return on investment

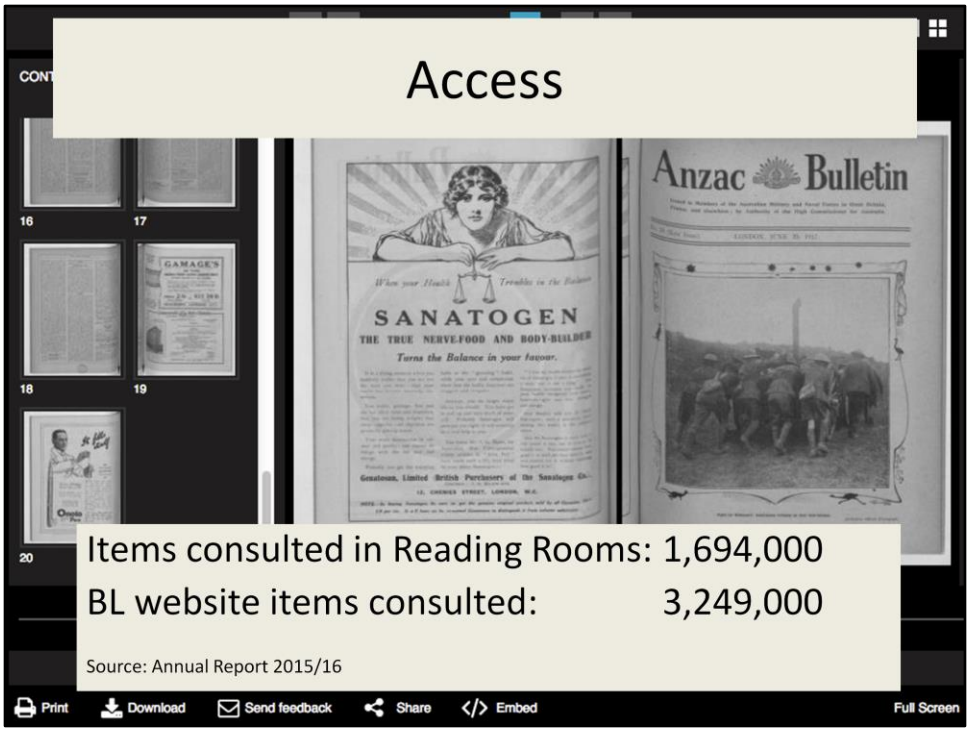
Why - what happens when you connect and open data?



Pragmatic - allows lots of people to use items without causing any further damage.
The storage void of the new British Library National Newspaper Building at Boston Spa in West Yorkshire. Photo © Kippa Matthews
TimeOut: 'Here, 60 million newspapers, spanning three centuries are kept in really neat low-oxygen storage chambers'



Digitised catalogue data is great but it's like hearing about a party you weren't at. Direct access to catalogue contents is even better. Digitisation is a key part of everyday business of cultural heritage institutions.



Why should access to our collective cultural and scientific heritage be limited to those who happen to be nearby, or who can afford to travel to see it? And why should it be limited to the opening hours of an organisation? (The BL reading rooms aren't open on a Sunday)

Why should access to our collective cultural and scientific heritage be limited to those who happen to be nearby, or who can afford to travel to see it? And why should it be limited to the opening hours of an organisation? (The BL reading rooms aren't open on a Sunday)

Lots of scholarship with digital collections ends up in traditional outputs, like monographs or articles. It can be incredibly difficult to track these uses, particularly if people cite the original and not the digital surrogate they actually used.

Delight

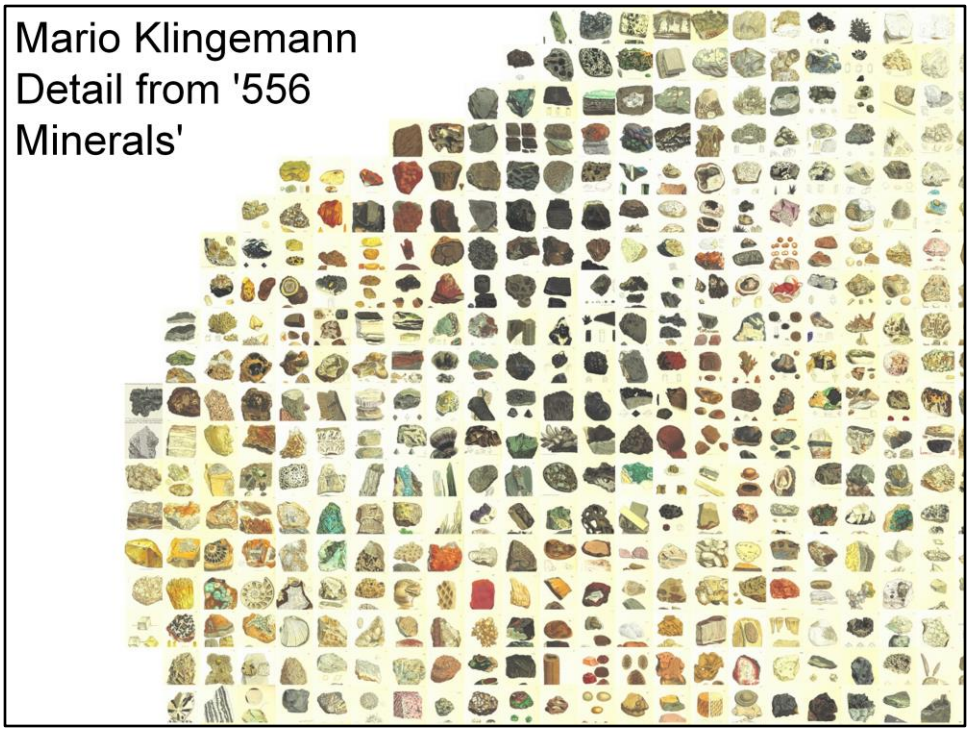
Some images have the ability to reach across centuries and delight us.

If you're a medieval scholar you might know the story that hedgehogs shake grape vines then 'trundle off back to their burrows, carrying the grapes on their spines, as a meal for their young', or the deeper moral about the devil, but you can delight in the image without that knowledge.

Cool things happen

Video clip for a Malaysian band with images from 19th century books shared on Flickr

This is a screenshot from a video clip made by a group from Malaysia using images from 19thC books the BL put on Flickr. They didn't need to ask us, but they were kind enough to email us afterwards.



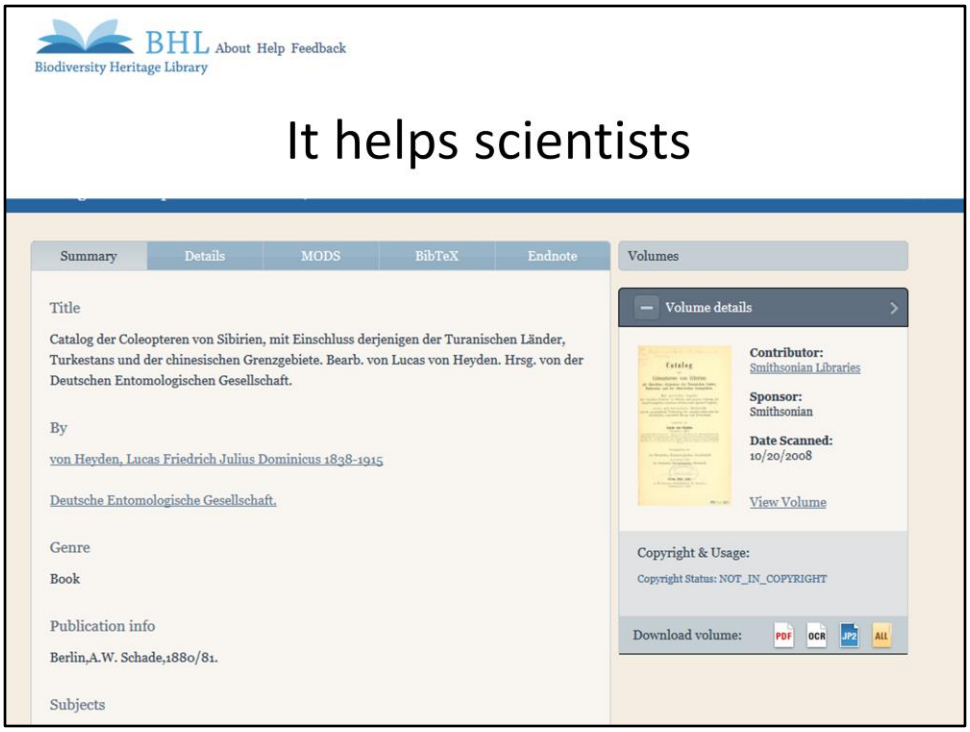
New ways of processing images as data, texture - the library could never have applied the technologies that code artist Mario Klingemann brought to 19th images. His exploration of the images resulted in new ways of seeing collections at scale.

His web site is: http://mario-klingemann.tumblr.com/

And http://incubator.quasimondo.com/

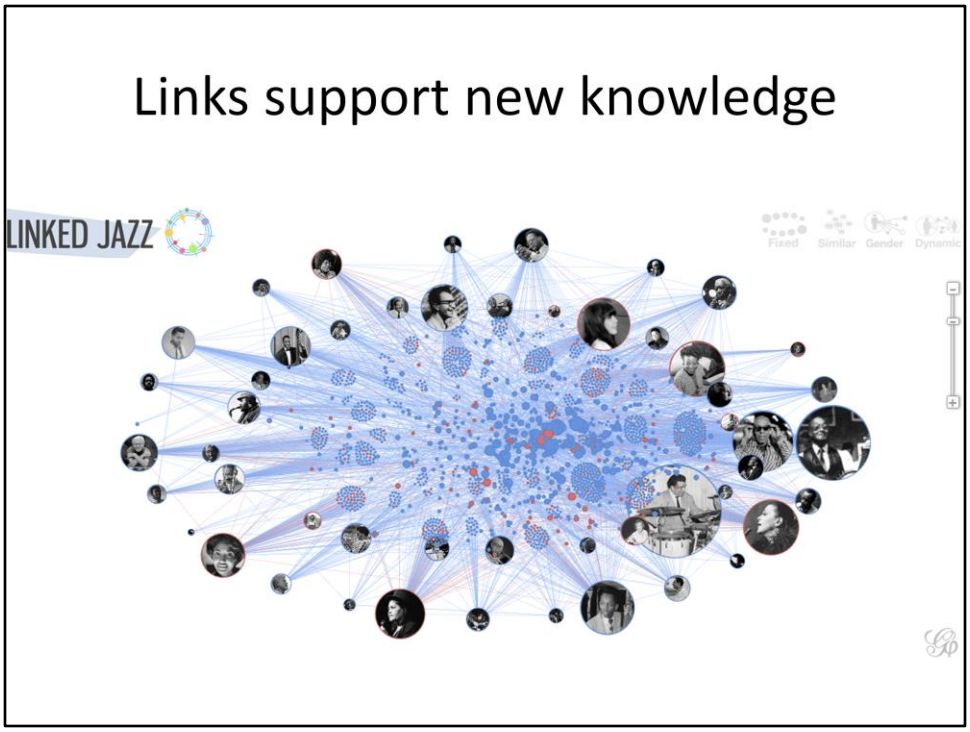
Partnerships without paperwork

When people can access and re-use digitised content, especially when it's identified unambiguously with links, they can do amazing things with it - without having to spend three years and lots of money on lawyers just to get started.



The Biodiversity Heritage Library improves research methodology by collaboratively making biodiversity literature openly available to the world as part of a global biodiversity community.

'These collections are of exceptional value because the domain of systematic biology depends, more than any other science, upon historic literature. Yet, this wealth of knowledge is available only to those few who can gain direct access to significant library collections. Literature about the biota existing in developing countries is often not available within their own borders.'



This site is able to pull in information about jazz musicians from a lot of other sites, making it much richer

https://linkedjazz.org/network/?mode=gender

Computational techniques applied to open data can help us learn more about collections. This project was able to identify woodcuts re-used in different publications and order the publications by finding tiny differences in the condition of the woodcuts.

Computer Vision and the History of Printing, Joon Son Chung

谢谢

Questions?

Open and connected datasets:
http://bit.ly/euWleR

Dr Mia Ridge @mia_out
Digital Curator, British Library

International Digital Culture Forum 2017
Taichung, Taiwan, August 2017