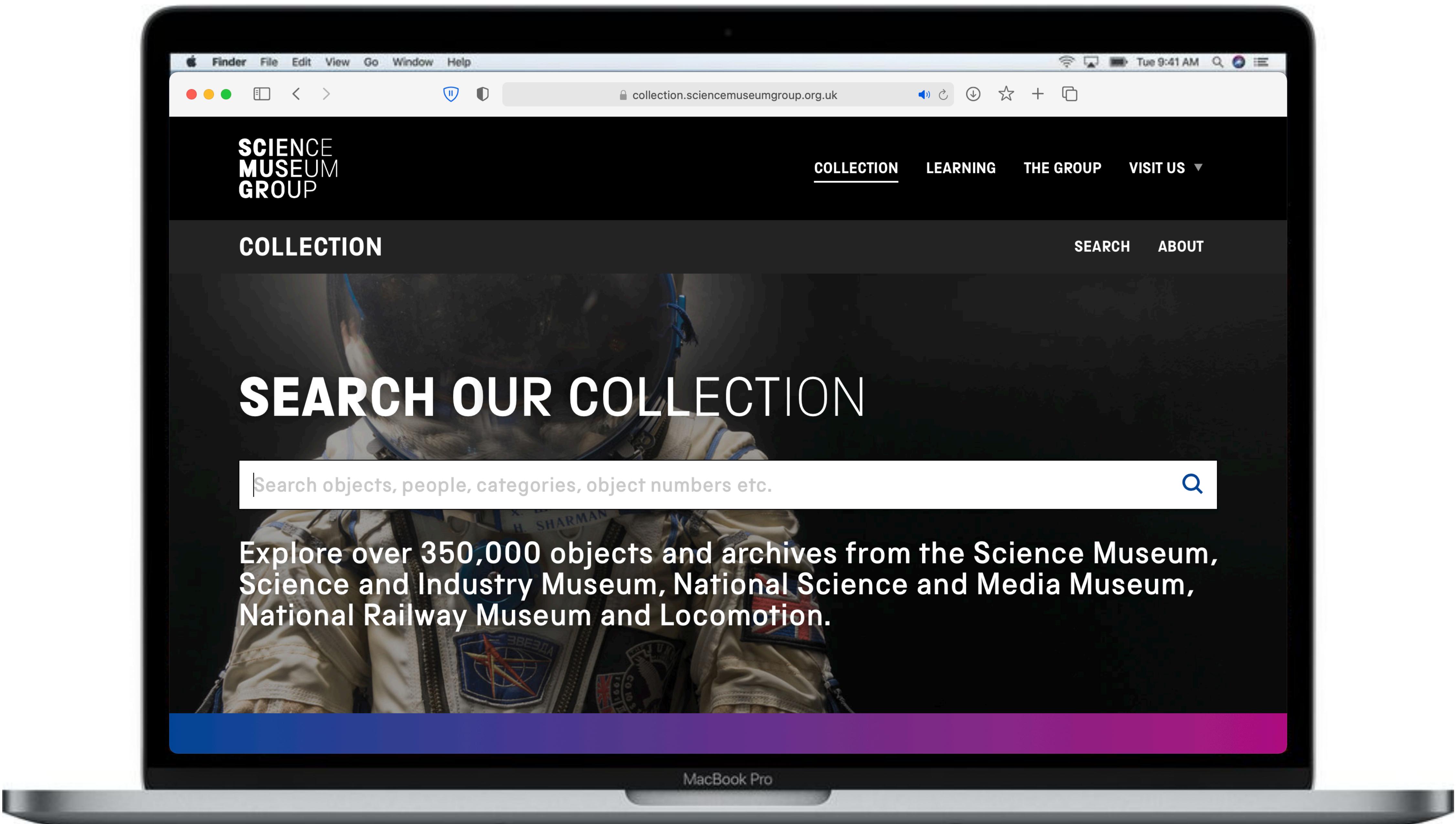


BUILDING LINKS WITH NATURAL LANGUAGE PROCESSING

2022 Smithsonian Digitisation Conference
John Stack, Digital Director
24 March 2022

SCIENCE
MUSEUM
GROUP





Finder File Edit View Go Window Help

collection.science museum group.org.uk/search/images?q=gwr

All 50 People 0 Objects 50 Documents 0

View:

Filter search

Geoffrey Tippett Collection - Swindon Works & GWR Locomotives [supplied title] Photographic Collections (Railway) c. 1930

THE GALVANIC AND ELECTRO-MAGNETIC TELEGRAPHHS,
ON THE
GT. WESTERN RAILWAY.
May be seen in constant operation, daily, (Sundays excepted) from 9 A.M. to 5 P.M., at the TELEGRAPH OFFICE, LONDON TERMINUS, PADINGTON AND TELEGRAPH COTTAGE, SLOUGH STATION.
An Exhibition adapted by its numerous Visitors to be the most interesting and attractive of any in this great Metropolis. In the list of visitors are the illustrious names of the nobility, gentry, and wealthier classes of Europe, and nearly the whole of the Nobility of England.
"This Exhibition, which has so much excited Public attention, is well worth seeing, and will be a great attraction to all classes of visitors."—Morning Post.
The Electric Telegraph is unlimited in the nature and extent of its communications; by its extraordinary agency a person in London could converse with another in New York, or with one in the far-off island of Tasmania, as rapidly as if both parties were in the same room. Questions proposed by Visitors will be asked by means of this Apparatus, and answers thereto will instantaneously be received from any part of the globe, without the intervention of a wire or a cable, in an incredibly short space of time, after the signal for his question has been given.
The Electric Fluid travels at the rate of 280,000 Miles per Second.
In the last year, 1,000,000 Messages have been sent, (as in the last case of Tawd,)—Telegrams detected; and lastly, which is of no little importance, the newly invented Medical aid has been procured in cases which otherwise would have proved fatal.

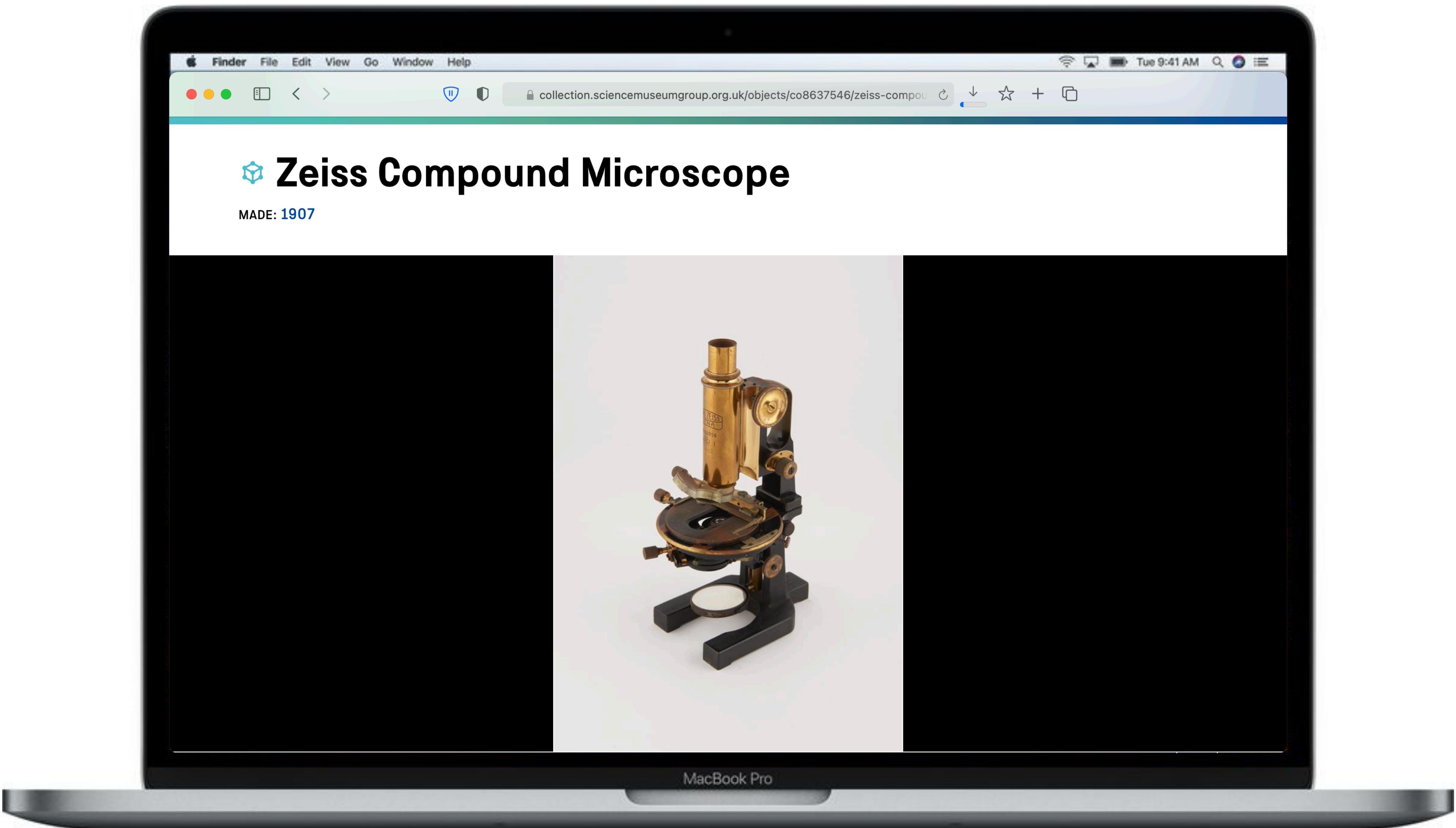
The great national importance of this wonderful invention is as well known than any further allusion here to its merits would be superfluous.
Mr. D. Dugdale, the author of the present communication, has stated that Messengers in instant attendance to that communication received by Telegraph, would bring so low a sum given.

The Electric Fluid travels at the rate of 280,000 Miles per Second.

1,000,000 Messages have been sent, (as in the last case of Tawd,)—Telegrams detected; and lastly, which is of no little importance, the newly invented Medical aid has been procured in cases which otherwise would have proved fatal.

Milk van, Great Western Railway Locomotives and Rolling Stock 1936

Diesel Railcar No 4, 1934, Great Western Railway Locomotives and Rolling Stock 1934



MacBook Pro

The image shows a MacBook Pro displaying a web browser window. The browser's title bar reads "collection.sciencemuseumgroup.org.uk/objects/co8637546/zeiss-compou". The main content area is titled "DETAILS" and contains the following information:

- CATEGORY: Scientific Instruments & Research
- OBJECT NUMBER: Y1991.49.2/1
- TYPE: compound microscope
- TAXONOMY:
 - furnishing and equipment
 - tools & equipment
 - optical instrument
 - microscope
 - furnishing and equipment
 - tools & equipment
 - optical instrument
 - microscope
 - CREDIT: Gift of Central Manchester Health Authority

To the right of the main content is a sidebar with the following sections:

 - CITE THIS PAGE**: A text input field containing "Science Museum Group. Zeiss Compound" with a copy icon.
 - RIGHTS**: Text stating "We encourage the use and reuse of our collection data." followed by two icons: one for Creative Commons Zero and one for Creative Commons Attribution 4.0 licence.
 - Using our data**: Links to download the catalogue entry as JSON, view an IIIF manifest in an IIIF viewer, add it to Animal Crossing Art Generator, and download the IIIF manifest.
 - DOWNLOAD**: A link to download the IIIF manifest.
 - A note at the bottom states: "Our records are constantly being enhanced and improved, but please note that we cannot guarantee the accuracy of any information shown on this website."

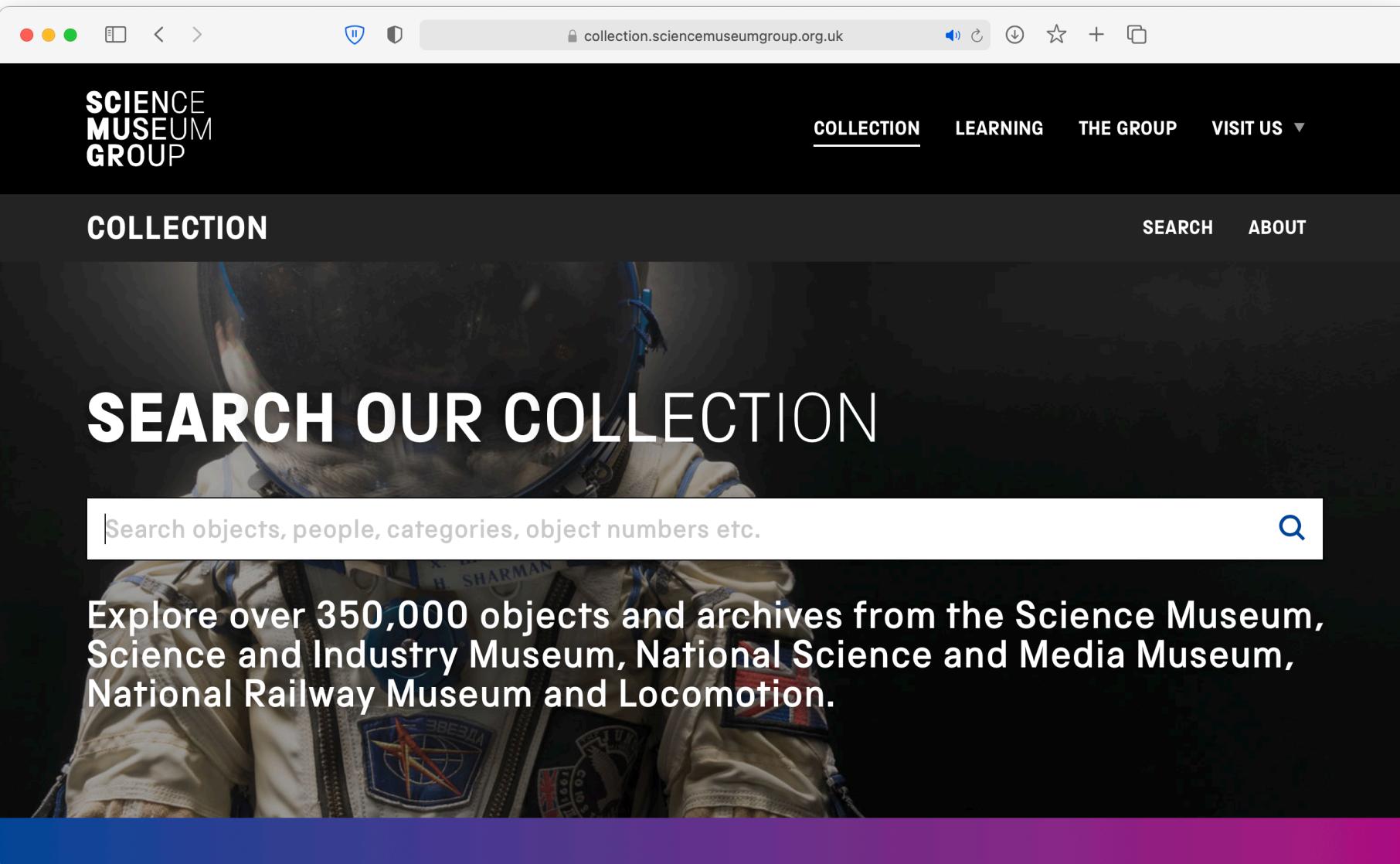
DISCOVERY AND EXPLORATION

1. Knowing about the collection
2. Understanding that it may contain relevant contents
3. Finding that content is available online
4. Navigating keyword-based search
5. Grasp the relevance of the content to the search
6. Navigate from that content to related content

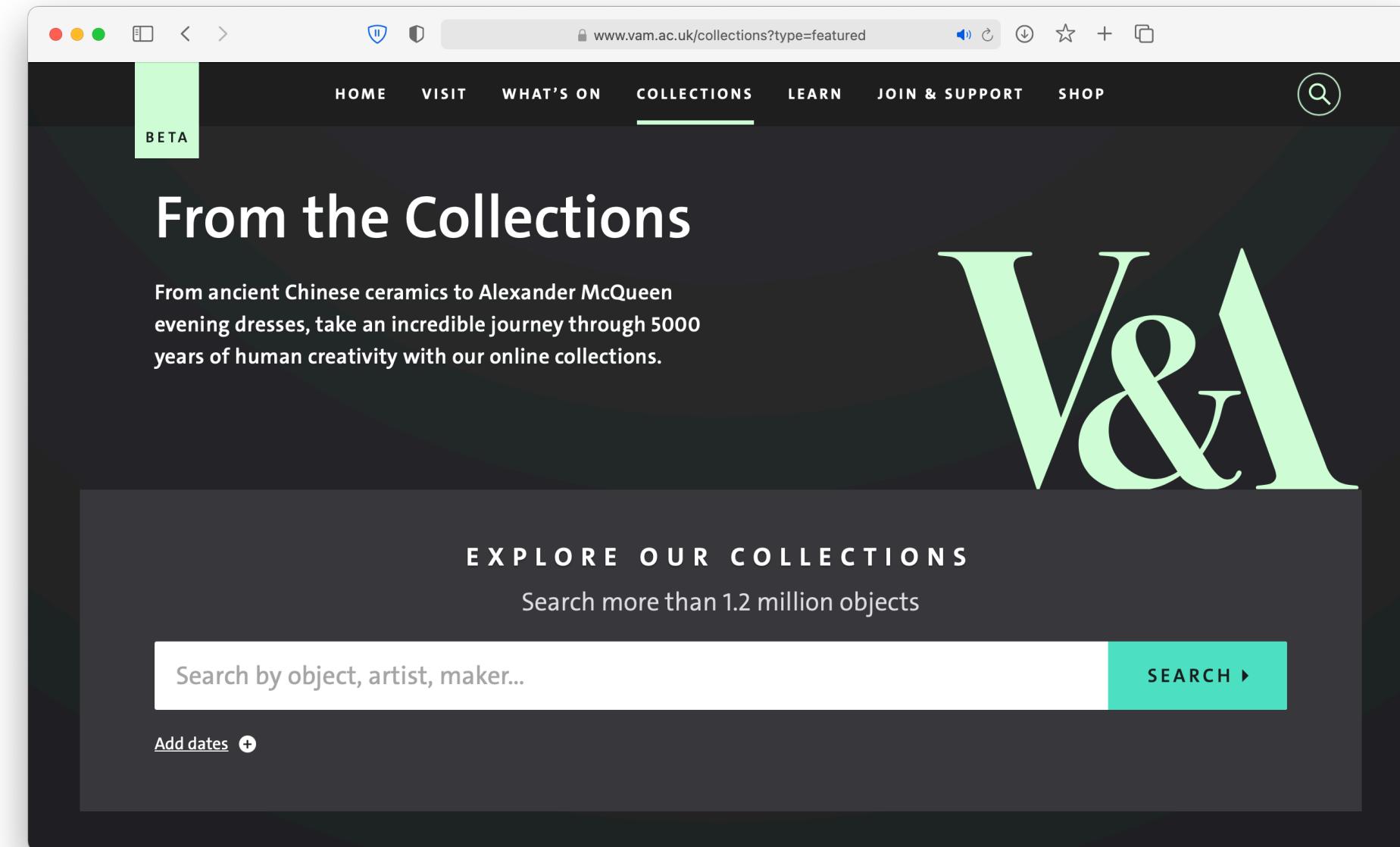
HERITAGE CONNECTOR PROJECT

How can existing digital tools and methods be used to build relationships at scale between poorly and inconsistently catalogued digitised collection objects and other content sources?

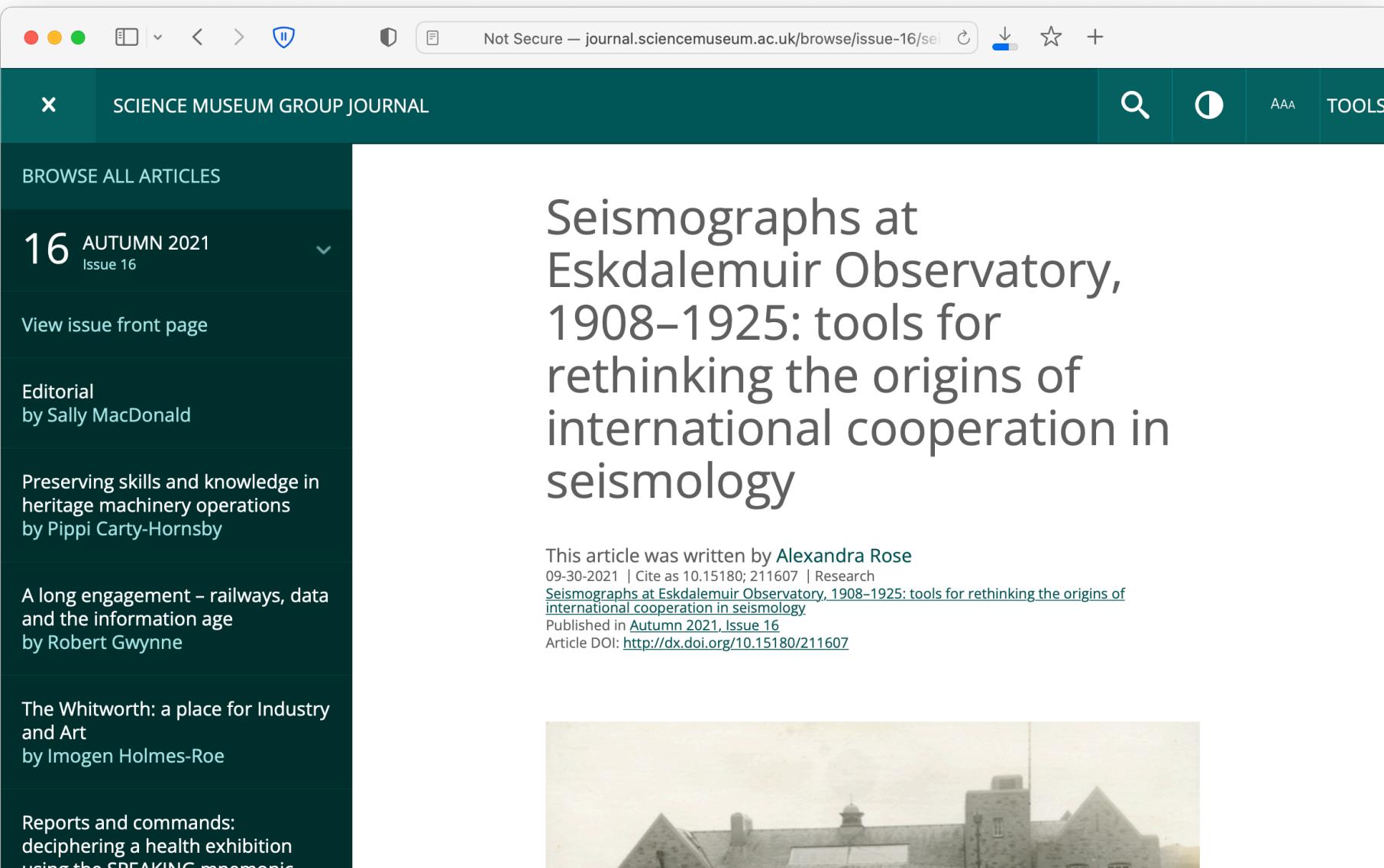
FIVE CONTENT SOURCES



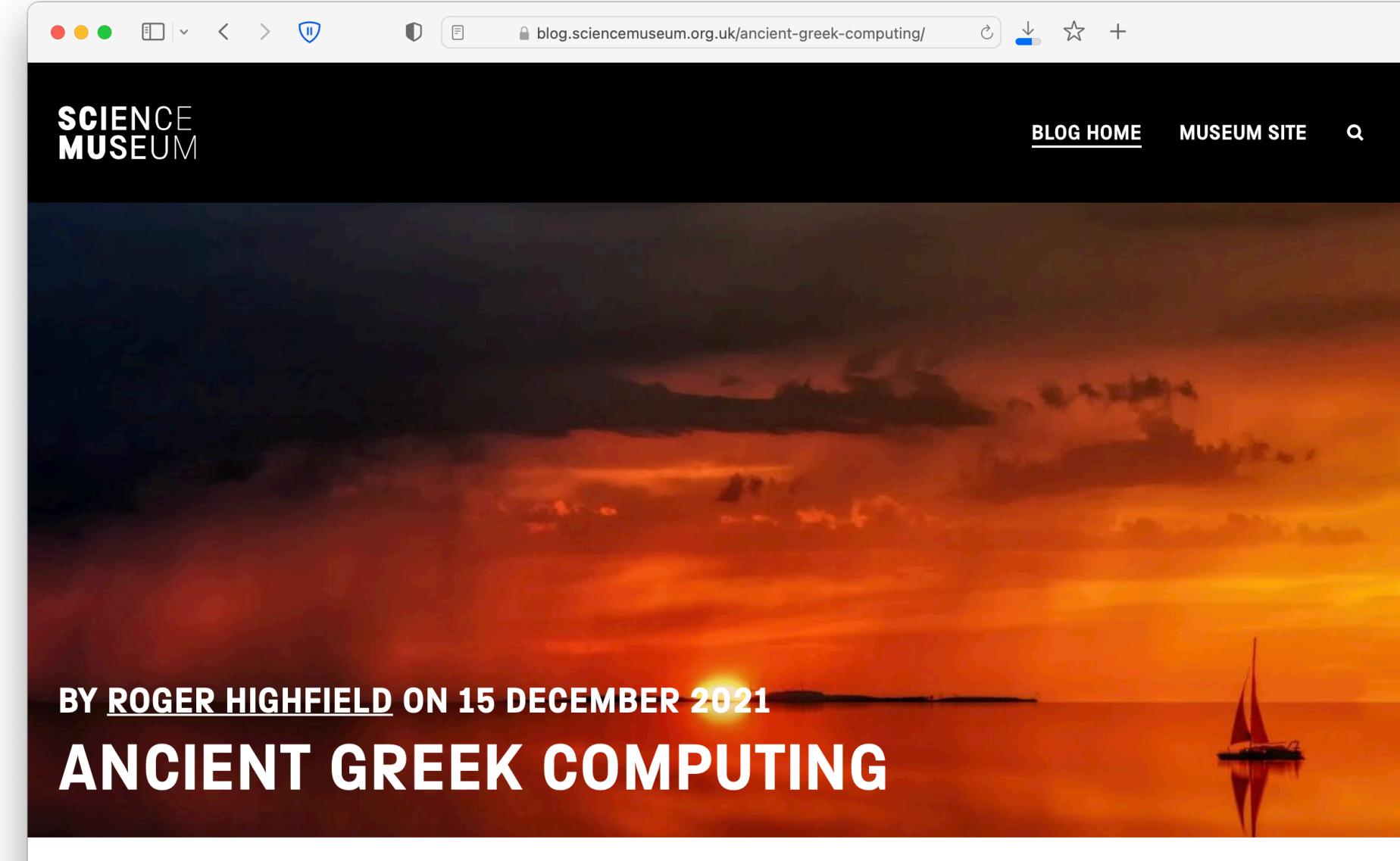
The screenshot shows the homepage of the Science Museum Group Collection website. At the top, there's a navigation bar with links for 'COLLECTION', 'LEARNING', 'THE GROUP', and 'VISIT US'. Below the navigation is a large banner with the text 'SEARCH OUR COLLECTION' and a search bar. A subtext below the banner reads: 'Explore over 350,000 objects and archives from the Science Museum, Science and Industry Museum, National Science and Media Museum, National Railway Museum and Locomotion.' The background of the main content area features a photograph of a宇航员 (astronaut) in a spacesuit.



The screenshot shows the homepage of the Victoria and Albert Museum (V&A) Collections website. It features a prominent green 'BETA' badge in the top left corner. The main heading is 'From the Collections', followed by a subtext: 'From ancient Chinese ceramics to Alexander McQueen evening dresses, take an incredible journey through 5000 years of human creativity with our online collections.' To the right is the V&A logo. Below the heading is a section titled 'EXPLORE OUR COLLECTIONS' with a search bar and a link to 'Search more than 1.2 million objects'.



The screenshot shows a journal article page from the Science Museum Group Journal. The left sidebar displays the journal's masthead (Issue 16, Autumn 2021), a table of contents, and various articles. The main content area features an article titled 'Seismographs at Eskdalemuir Observatory, 1908–1925: tools for rethinking the origins of international cooperation in seismology' by Alexandra Rose. The article includes a small thumbnail image of a building and some descriptive text at the bottom.



The screenshot shows a blog post from the Science Museum. The header features the Science Museum logo and navigation links for 'BLOG HOME', 'MUSEUM SITE', and a search icon. The main image is a dramatic sunset or sunrise over water. The title of the post is 'ANCIENT GREEK COMPUTING' and it is attributed to 'BY ROGER HIGHFIELD ON 15 DECEMBER 2021'.

en.wikipedia.org/wiki/LNER_Class_A3_4472_Flying_Scotsman

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

LNER Class A3 4472 Flying Scotsman

From Wikipedia, the free encyclopedia

This article is about the locomotive. For the train service, see [Flying Scotsman \(train\)](#). For other uses, see [Flying Scotsman \(disambiguation\)](#).

LNER Class A3 4472 Flying Scotsman is a Pacific steam locomotive built in 1923 for the London and North Eastern Railway (LNER) at Doncaster Works to a design of Nigel Gresley. It was employed on long-distance express [East Coast Main Line](#) trains by the LNER and its successors, British Railways Eastern and North Eastern Regions, notably on the [London to Edinburgh Flying Scotsman](#) train service after which it was named.

The locomotive set two world records for steam traction, becoming the first steam locomotive to be officially authenticated at reaching 100 miles per hour (160.9 km/h) on 30 November 1934.^[1] and then setting a record for the longest non-stop run by a steam locomotive when it ran 422 miles (679 km) on 8 August 1989 while in Australia.^[2]

Retired from regular service in 1963 after covering 2.08 million miles,^{[1][3][4]} Flying Scotsman enjoyed considerable fame in preservation under the ownership of, successively, Alan Pegler, William McAlpine, Tony Marchington, and finally the [National Railway Museum](#) (NRM).

As well as hauling enthusiast specials in the United Kingdom, the locomotive toured extensively in the United States and Canada from 1969 until 1973^[5] and Australia in 1988/89.^[6] Flying Scotsman has been described as the world's most famous steam locomotive.^{[7][8]} In a 2015 poll which questioned people from four continents it was again ranked the most famous locomotive.^[9]

Contents [hide]

- 1 History
- 2 Preservation
 - 2.1 Alan Pegler
 - 2.2 William McAlpine
 - 2.3 Tony Marchington
 - 2.4 National Railway Museum
 - 2.5 Overhaul 2006–2016

Flying Scotsman

Flying Scotsman on the West Somerset Railway on 11 September 2017 in BR green livery with prominent German-style smoke deflectors and double chimney.

Type and origin	[hide]
Power type	Steam
Designer	Sir Nigel Gresley
Builder	Doncaster Works
Build date	February 1923
Website	www.flyingscotsman.org.uk

Specifications

Configuration:	
• Whyte	4-6-2
Gauge	4 ft 8½ in (1,435 mm) standard gauge
Driver dia.	80 in (2,032 mm) diameter



Finder File Edit View Go Window Help

www.wikidata.org/wiki/Q54173

English Not logged in Talk Contributions Create account Log in

Item Discussion Read View history Search Wikidata

General Electric (Q54173)

American multinational conglomerate corporation

General Electric Company | GE | General Electric Co.

In more languages

Language	Label	Description	Also known as
English	General Electric	American multinational conglomerate corporation	General Electric Company GE General Electric Co.
British English	General Electric	No description defined	
French	General Electric	conglomérat américain	
German	General Electric	US-amerikanischer Mischkonzern	

All entered languages

Statements

instance of	business	edit
	▼ 0 references	+ add reference
	enterprise	edit
	▼ 0 references	

MacBook Pro



✿ Marks Tey Station With Mr L Sydney Bracher

MADE: 1899-06-06 in Marks Tey Railway Station

MAKER: Edgar Tarry Adams



'Marks Tey Station With Mr L. Sydney Bracher, 1899 A photograph of Mr L
Science Museum Group Collection
© The Board of Trustees of the Science Museum

↗ | > Use this image

A photograph entitled 'Marks Tey Station With Mr L. Sydney Bracher', taken by Edgar Tarry Adams in June 1899. Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

A photograph entitled 'Marks Tey Station With Mr L Sydney Bracher', taken by Edgar Tarry Adams in June 1899.

Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

DETAILS

CATEGORY: [Photographs](#)

OBJECT NUMBER: 1997-5003/157

TYPE: [photograph](#)

TAXONOMY: visual and verbal communication

CITE THIS PAGE

Science Museum Group. Marks Tey Station



RIGHTS

We encourage the use and reuse of our collection data.



Data in the title, made, maker and details fields are released under [Creative Commons](#)



_marks Tey Station With Mr L Sydney Bracher

MADE: 1899-06-06 in Marks Tey Railway Station

MAKER: Edgar Tarry Adams



'Marks Tey Station With Mr. L. Sydney Bracher, 1899 A photograph of Mr L. Science Museum Group Collection
© The Board of Trustees of the Science Museum



Use this image

A photograph entitled 'Marks Tey Station With Mr L. Sydney Bracher', taken by Edgar Tarry Adams in June 1899. Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

A photograph entitled 'Marks Tey Station With Mr L Sydney Bracher', taken by Edgar Tarry Adams in June 1899.

Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

DETAILS

CATEGORY: Photographs

OBJECT NUMBER: 1997-5003/157

TYPE: photograph

TAXONOMY: visual and verbal communication

CITE THIS PAGE

Science Museum Group. Marks Tey Station

RIGHTS

We encourage the use and reuse of our collection data.

the title, made, maker and details fields are released under Creative Commons

SCIENCE MUSEUM GROUP

COLLECTION SEARCH ABOUT VISIT OUR MUSEUMS ▾

Search objects, people, categories, object numbers etc.

Marks Tey Station With Mr L Sydney Bracher

MADE: 1899-06-06 in Marks Tey Railway Station

MAKER: Edgar Tarry Adams



'Marks Tey Station With Mr. L. Sydney Bracher, 1899 A photograph of Mr L. Science Museum Group Collection
© The Board of Trustees of the Science Museum

A photograph entitled 'Marks Tey Station With Mr L. Sydney Bracher', taken by Edgar Tarry Adams in June 1899. Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

A photograph entitled 'Marks Tey Station With Mr L Sydney Bracher', taken by Edgar Tarry Adams in June 1899.

Mr Bracher, probably a friend of Adams, is captured smoking his pipe and tipping his straw boater at the camera at the station in Essex.

DETAILS

CATEGORY: Photographs

OBJECT NUMBER: 1997-5003/157

TYPE: photograph

TAXONOMY: visual and verbal communication

CITE THIS PAGE

Science Museum Group. Marks Tey Station

RIGHTS

We encourage the use and reuse of our collection data.

Data in the title, made, maker and details fields are released under Creative Commons



Item Discussion

English Not logged in Talk Contributions Create account Log in

Read View history Search Wikidata

Marks Tey railway station (Q2025932)

No description defined

In more languages

Statements

instance of	railway station	edit
	▼ 0 references	+ add reference
	+ add value	

image		edit
	Markstey1.jpg	648 × 429; 98 KB
	► 1 reference	+ add value

named after	Marks Tey	edit
	▼ 0 references	+ add reference
	+ add value	

country	United Kingdom	edit
	▼ 0 references	+ add reference
	+ add value	

located in the administrative territorial entity	Marks Tey	edit
	► 1 reference	+ add value

coordinate location		edit
	51°52'51"N, 0°46'58"E	Wikimedia maps Map data © OpenStreetMap
	► 1 reference	+ add value

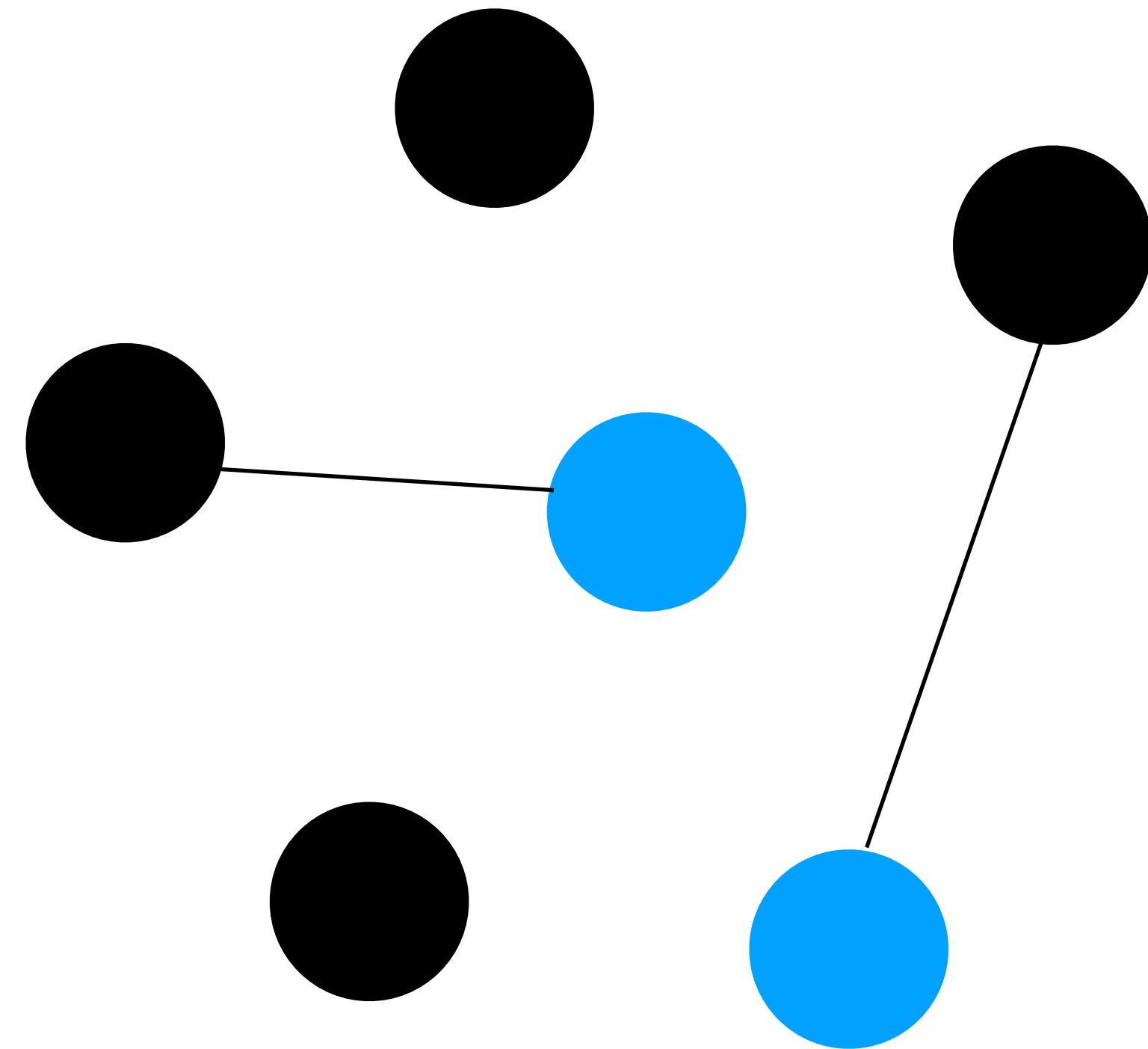
operator	Greater Anglia	edit



LINK BUILDING

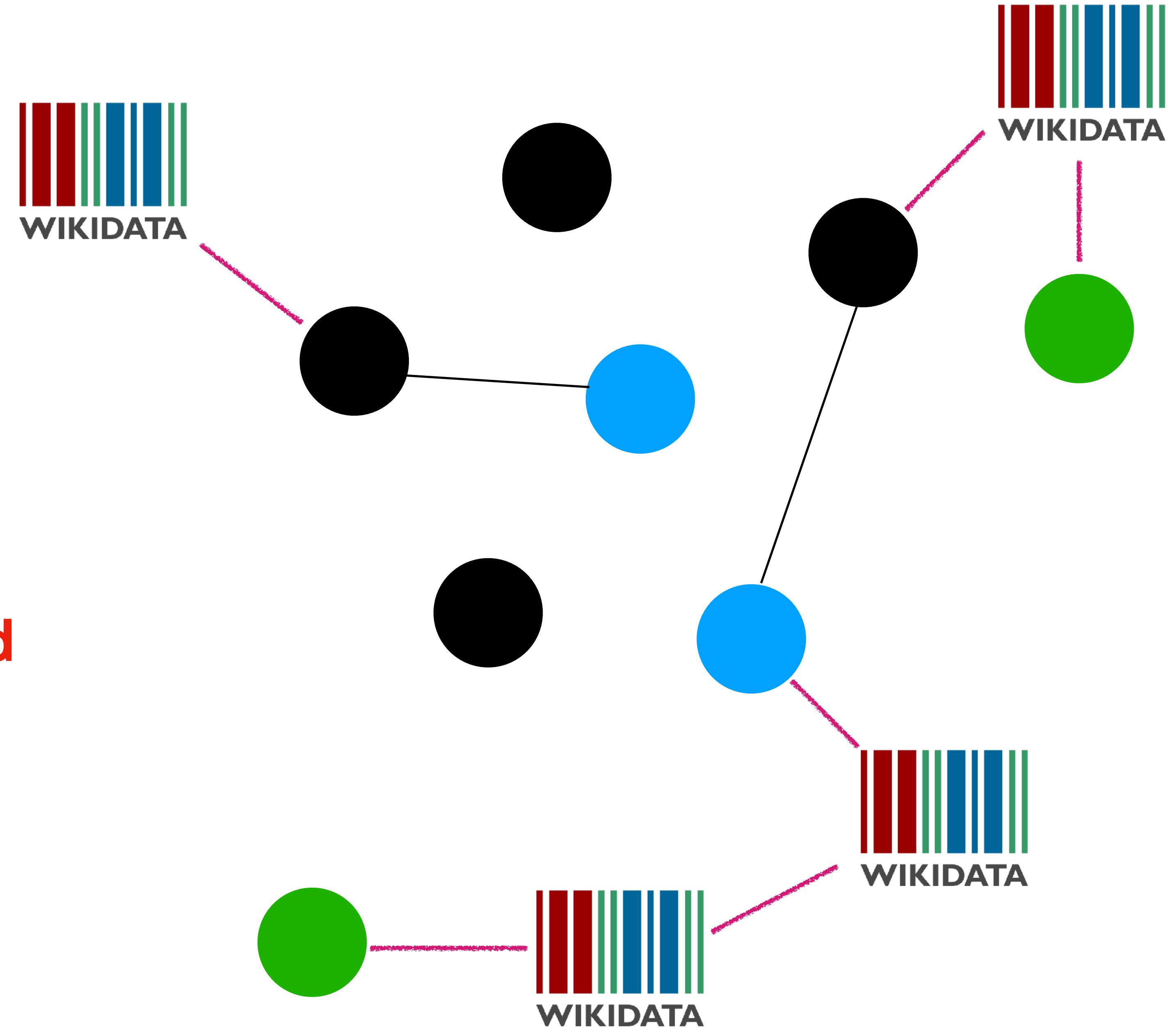
**THIS IS OUR
COLLECTION
NOW..**

Small islands of thin data



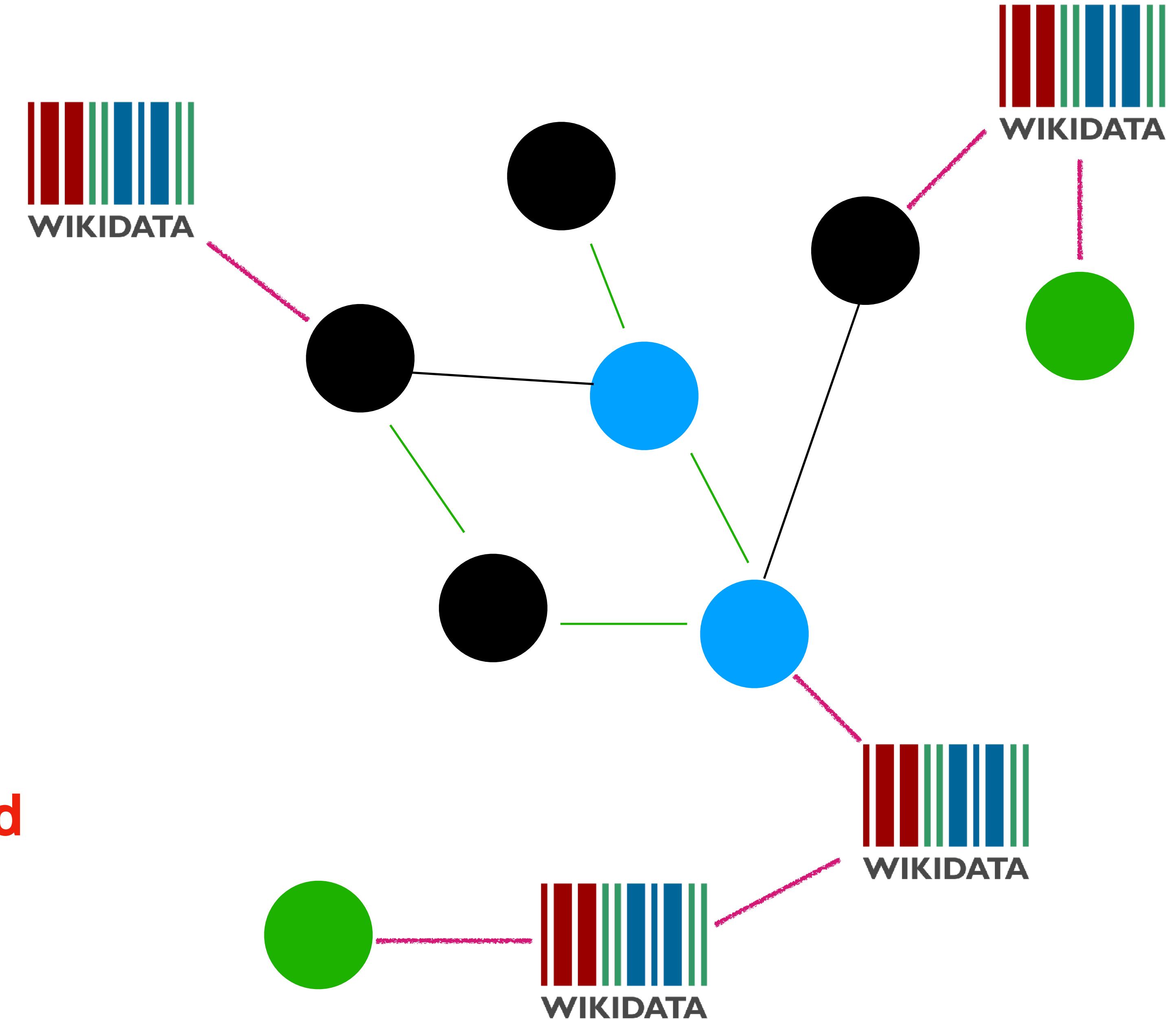
**THIS IS OUR
COLLECTION
CONNECTED TO
WIKIDATA..**

**Small islands of connected
data**



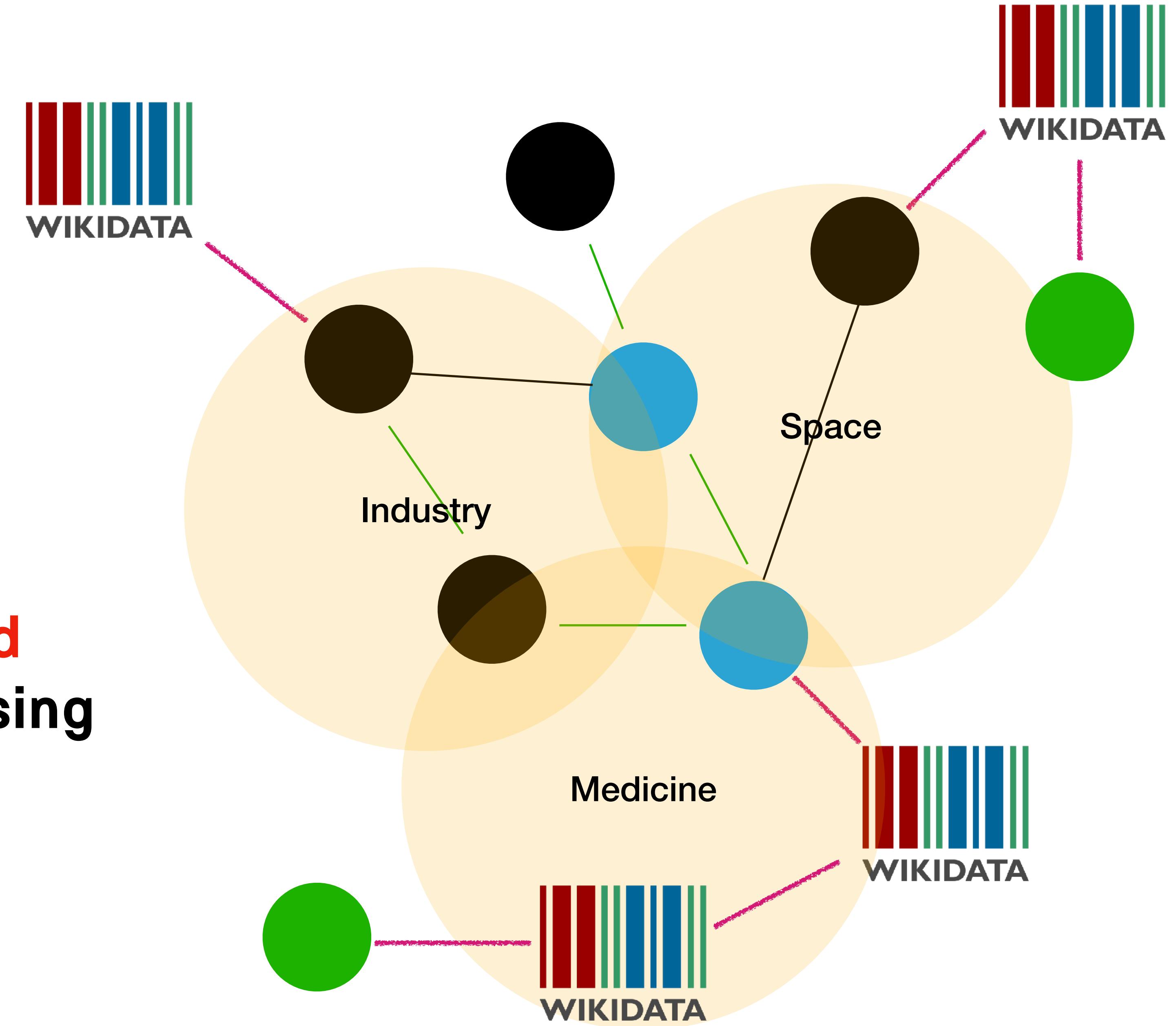
**THIS IS OUR
COLLECTION
INTERLINKED VIA
INFORMATION
EXTRACTION
TECHNIQUES..**

**Small islands of connected
and interlinked data**



THIS IS OUR COLLECTION WITH NEW GROUPINGS..

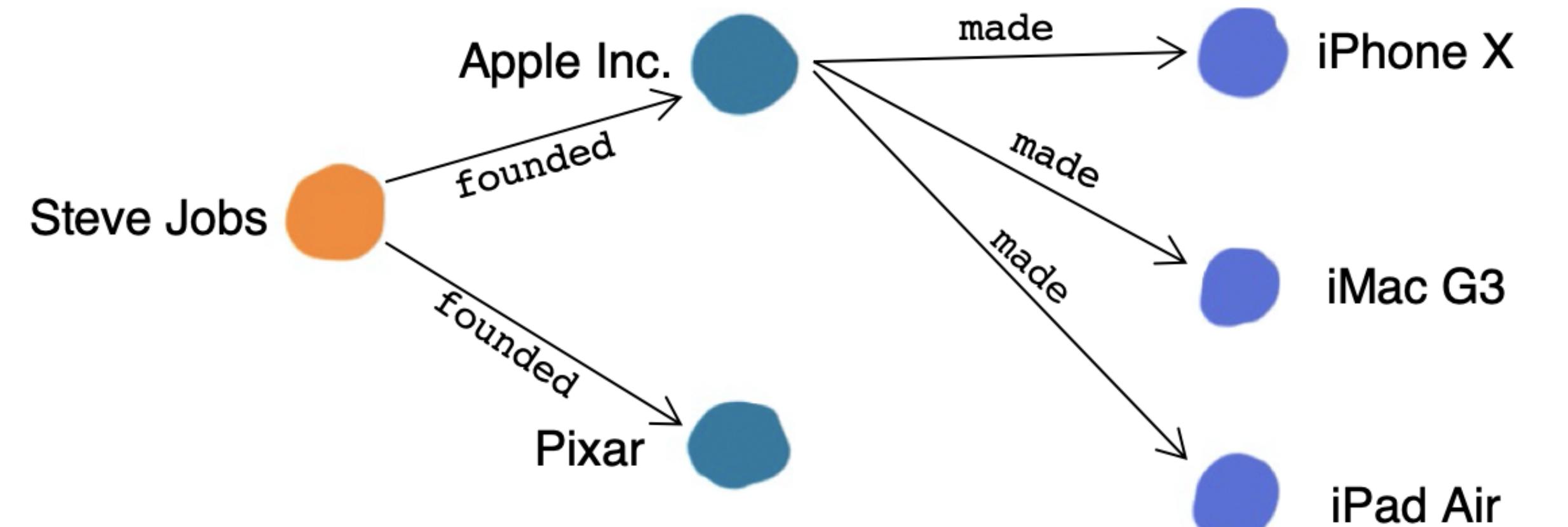
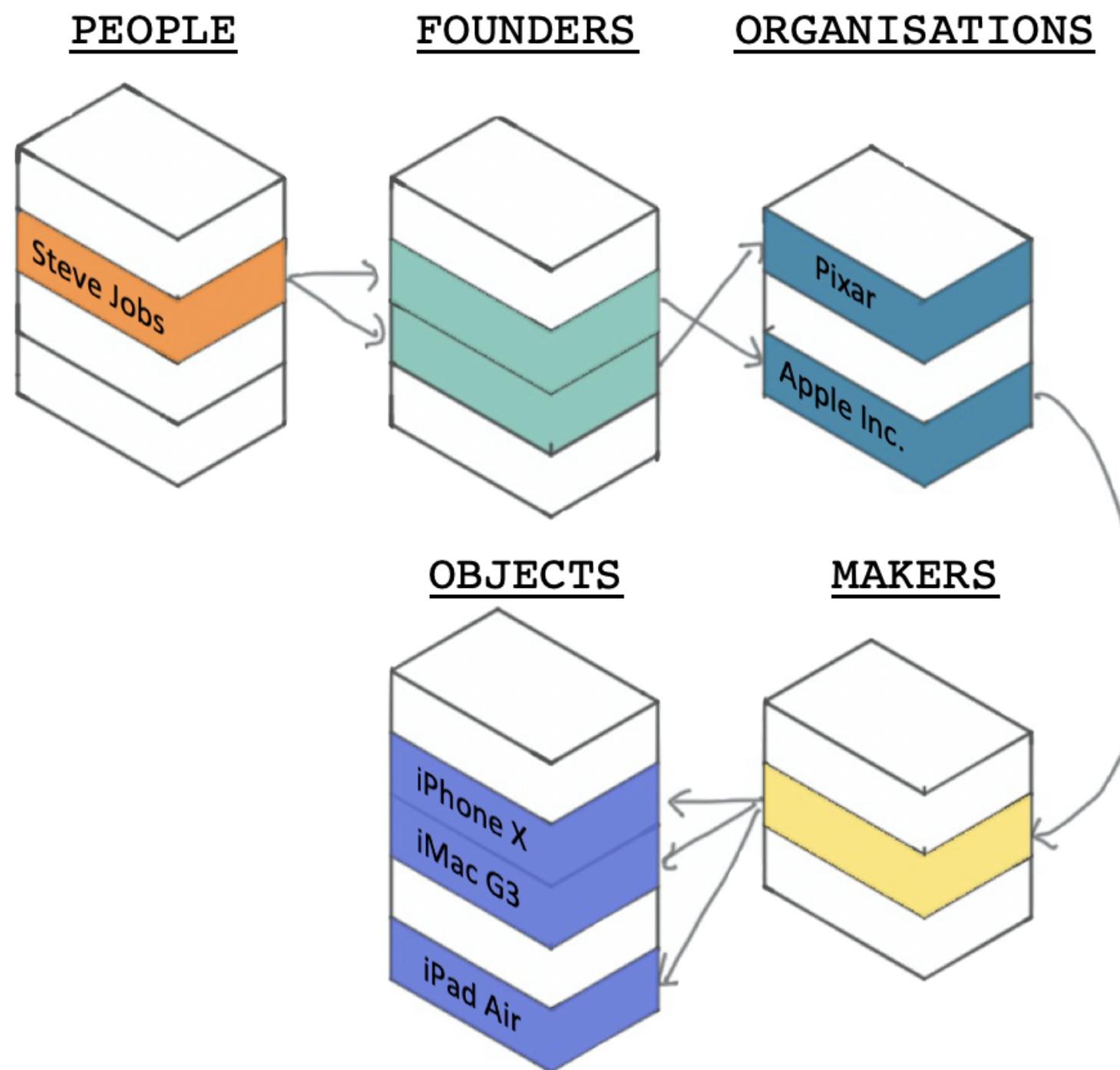
Small islands of **connected**
and **interlinked** data exposing
new **groupings**



THREE TECHNOLOGIES

1. Machine Learning (specifically Natural Language Processing)
2. Linked data
3. Knowledge graphs

LINKED DATA AND KNOWLEDGE GRAPHS



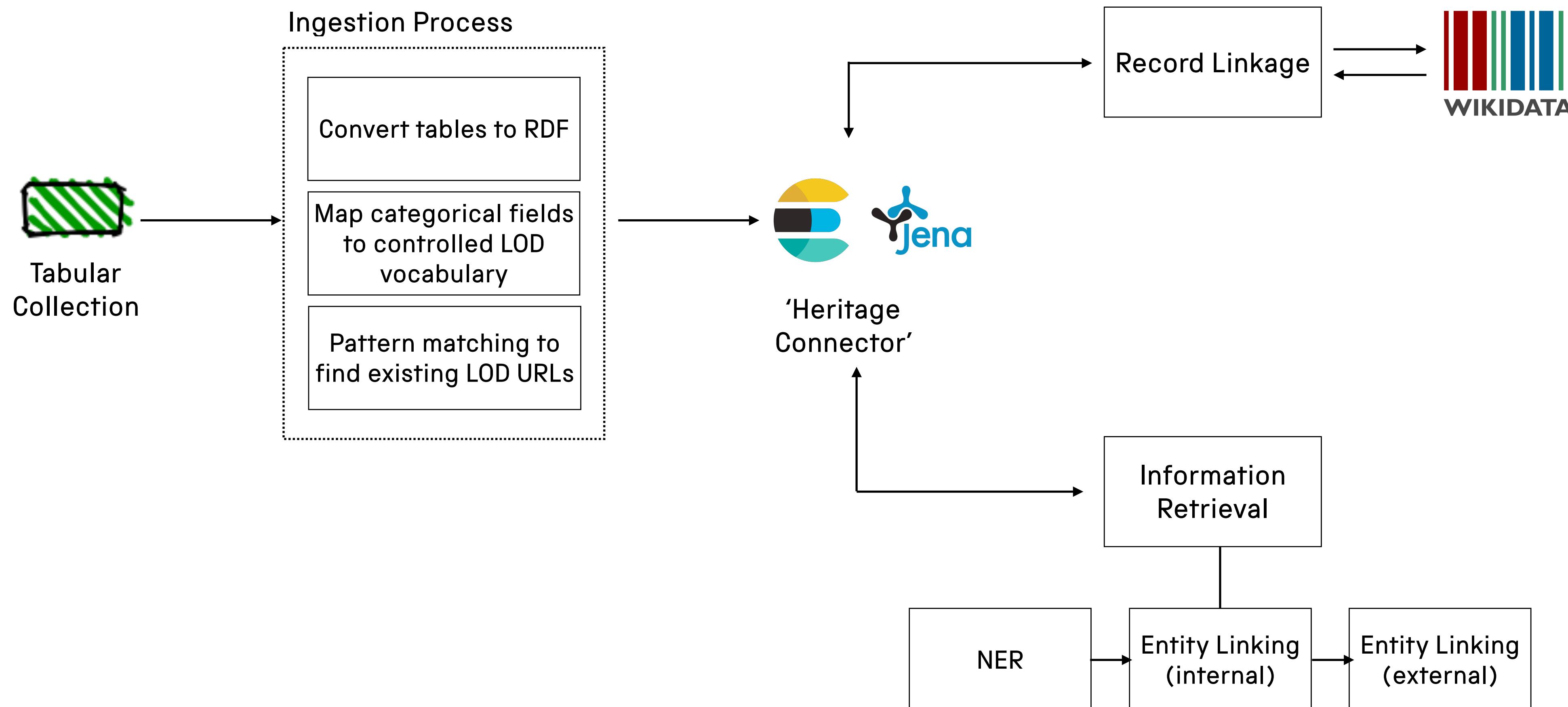
text Anna Atkins was born in Tonbridge, Kent

triples anna_atkins, birth_place, tonbridge_kent

linked data

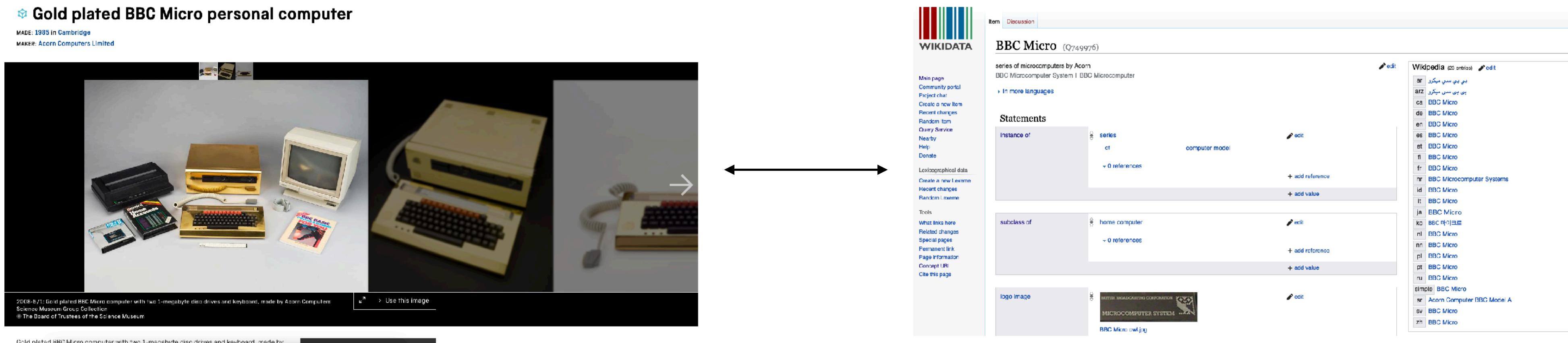
<https://collection.science museum group.org.uk/people/cp11399>,
<https://www.wikidata.org/wiki/Property:P19>,
<https://www.wikidata.org/wiki/Q936183>

BUILDING THE KNOWLEDGE GRAPH: TECHNICAL SUMMARY

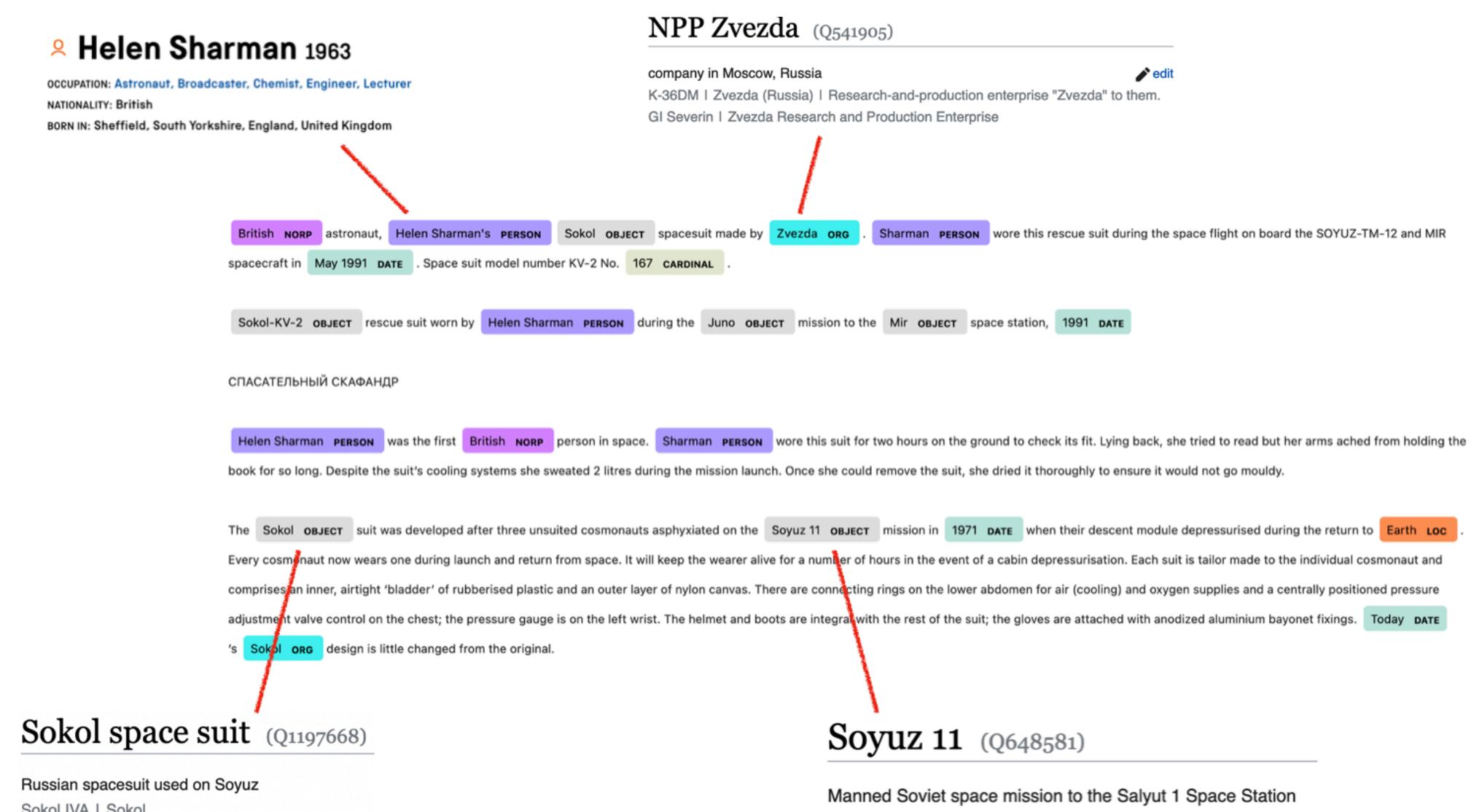


NATURAL LANGUAGE PROCESSING

1. Record Linkage with Wikidata (-> Wikipedia, other collections and databases)



2. Named Entity Recognition (NER) and Entity Linking (EL)



SOME HIGH-LEVEL STATISTICS

- ~354,000 Science Museum Group collection, blog and journal pages and ~325,000 V&A collection records have been processed.
- ~8,700 object, person and organisation have direct connections to Wikidata (only Science Museum Group processed).
- Using NER and EL:
 - ~484,000 connections have been added between collection records,
 - ~648,000 connections have been added to Wikidata,
 - ~1.2 million connections have been added to yet unknown entities.
- In total, there are ~110,000 Wikidata/Wikipedia pages in the knowledge graph.

USER INTERFACES

AFFORDANCES OF THIS NEW DATASET

- **macro-views** of the entire collections (plus the blog and journal).
- **serendipitous discovery**, through linking of previously isolated records.
- **new entry points** into the collection from Wikipedia and Wikidata (events, bands, etc.).
- **new forms of interface**, to provide an alternative to keyword search.

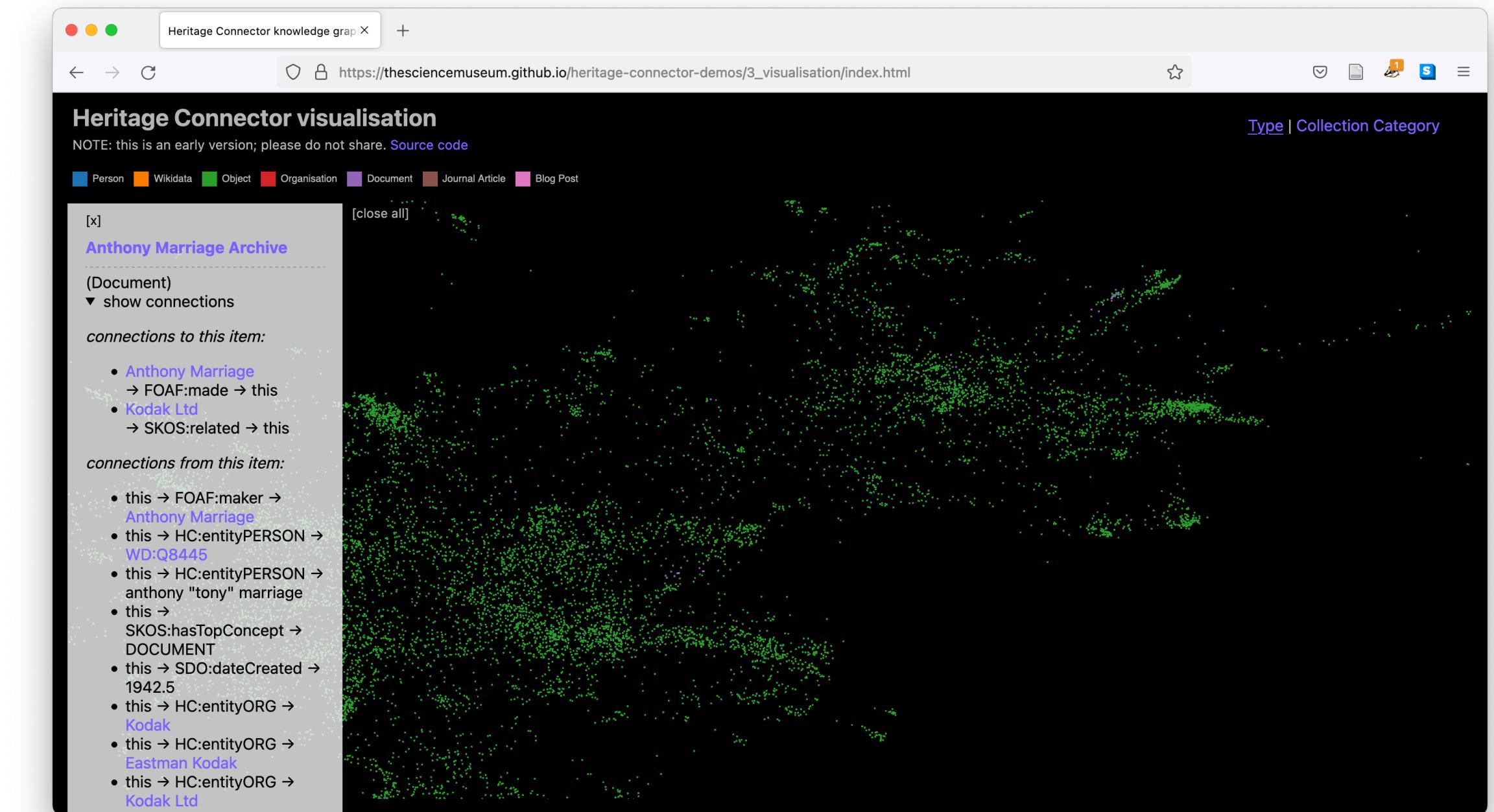
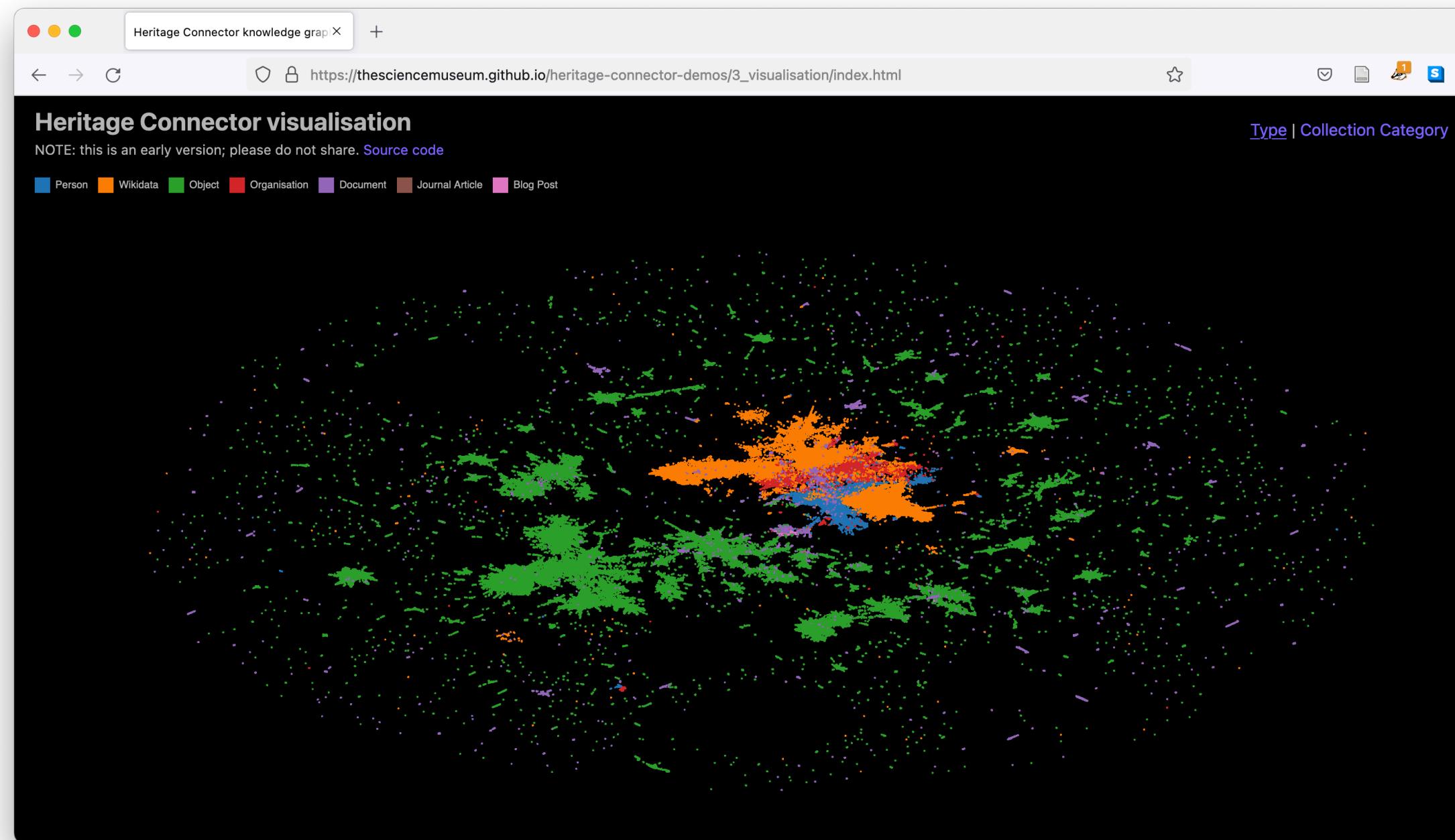
DEMONSTRATORS

This page serves as a directory of demos for the [Heritage Connector project](#). The source code for most of these demos can be found on [Github](#).

The screenshot shows a web browser window with the title "Heritage Connector Demonstrators". The page content includes:

- Mapping locations:** A map of the world with numerous blue location pins.
- Metadata explorer:** A search interface for exploring the knowledge graph. It includes a sidebar with "Project info" and a search bar with placeholder "...or enter a URL". Below the search bar is a note: "This URL can be from the Science Museum Group's online collections, blog or academic journal, or from some select categories in the world, as we processed our first batch of data in 2020."
- Visualisation:** A network graph visualization showing connections between entities. Nodes are colored by category, such as orange for Collection, green for Person, and blue for Organization.
- Visualisation - multiple collections:** A similar network graph visualization, but with more nodes and a denser cluster of connections compared to the single collection visualization.
- Prototype four track tape recorder for Oramics Machine:** A detailed object page for a specific item in the collection.
- Heritage Connector SPARQL endpoint:** A screenshot of the SPARQL query editor showing a query and its results as a network graph.

MACRO COLLECTION VISUALISATION



BOOKMARKLET

collection.science museum group.org.uk/objects/co226606/spirit-of-the-southern

SCIENCE MUSEUM GROUP

COLLECTION

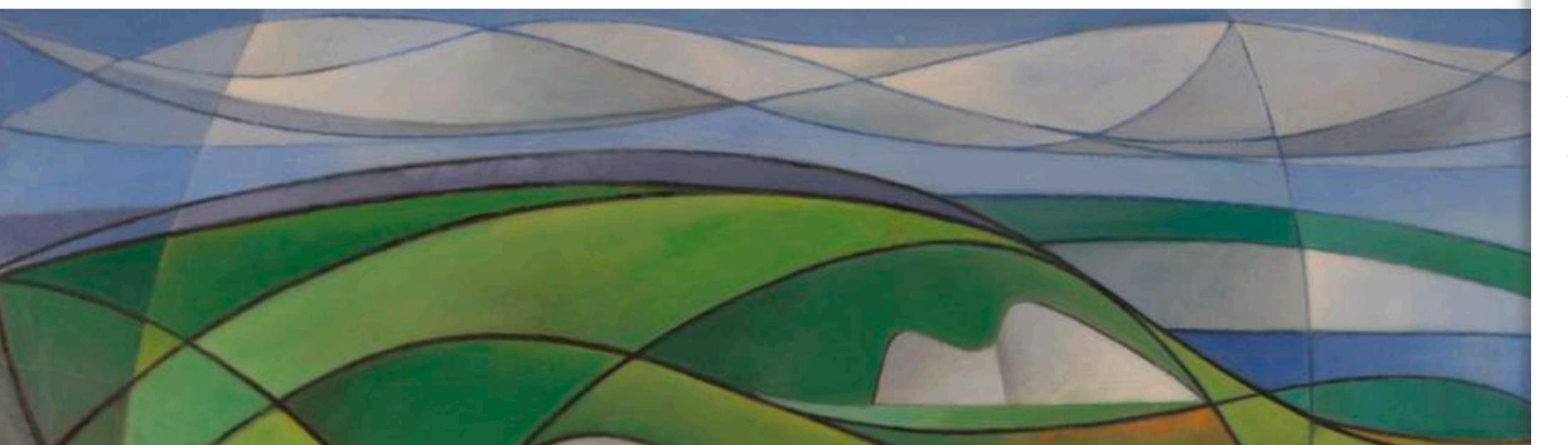
Search objects, people, categories, object numbers etc.

Spirit of the Southern

MADE: 1969, about in England

COMMISSIONER: British Railways, Southern Region

ARTIST: Harry Stevens



Heritage Connector - connections found
Log interesting/incorrect connections (HC team)

EXISTING RECORD METADATA

- RDFS:label
 - Spirit of the Southern
- HC:database
 - mimsy
- RDF:type
 - WD:Q3305213
- SKOS:hasTopConcept
 - OBJECT
- SDO:isPartOf
 - Pictorial Collection (Railway)
- XSD:additionalType
 - painting
- SDO:material
 - acrylic paint
- SDO:dateCreated
 - 1969.0

EXISTING LINKS TO OTHER PAGES

FOAF:maker

- British Railways, Southern Region [SMGP:cp44102]
- Harry Stevens [SMGP:cp38391]

SAMEAS/SIMILAR LINKS

FROM NER & NEL

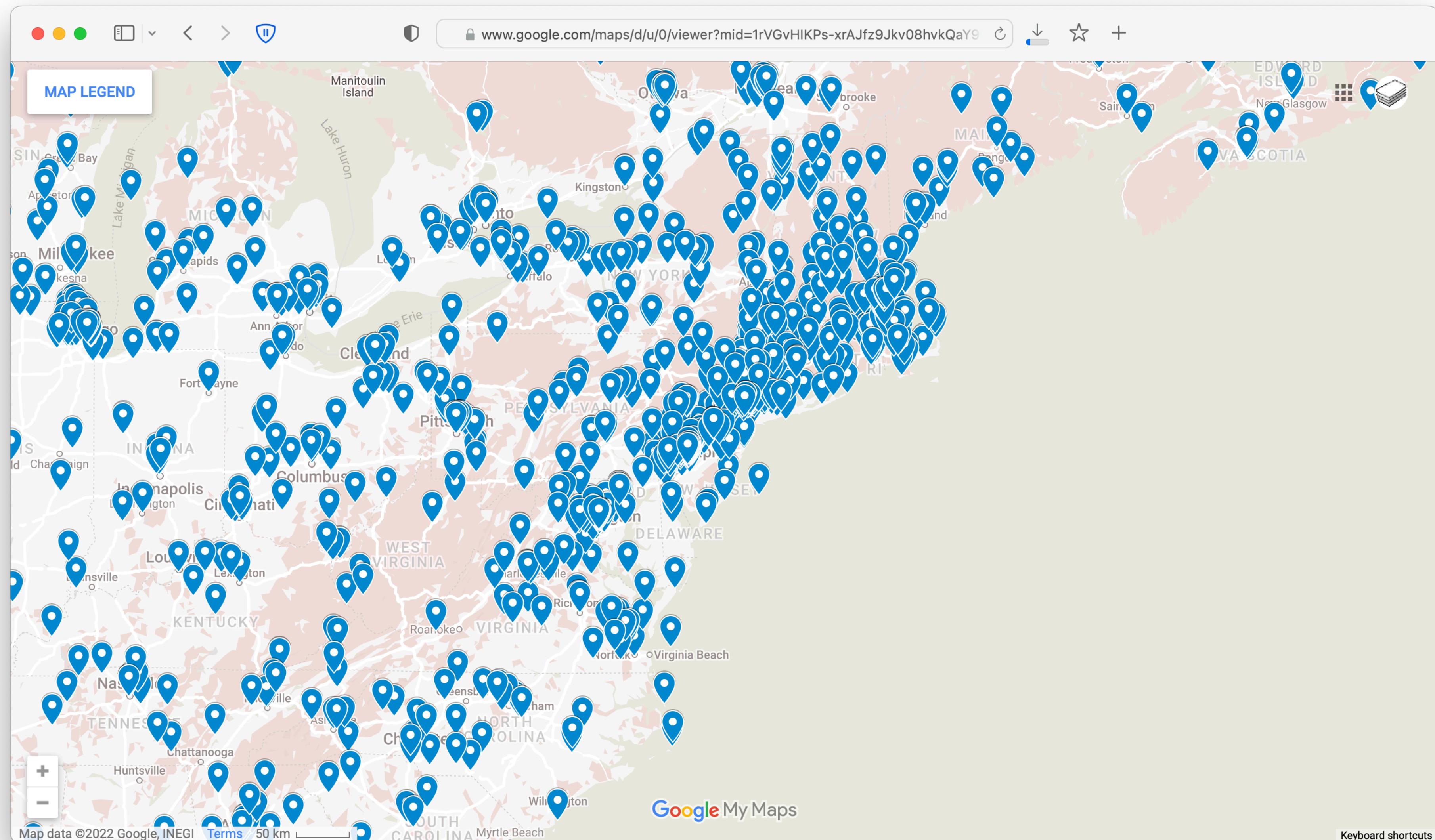
HC:entityPERSON

- I.w ibbotson
- Harry Stevens [SMGP:cp38391]

HC:entityORG

- British Railways [SMGP:cp3862]
- British Railways [SMGP:ap315]
- Southern Region of British Railways [WD:Q7570287]

MAP



FINDINGS...

- Methods used can generate huge numbers of links.
- Fruitful avenues for improved search and discovery.
- Collection text fields and article content are valuable locations for link building.
- Opportunities in new data alongside collection catalogue.
- False positive results are usually readily apparent.
- New forms of interface can be developed (and are needed).

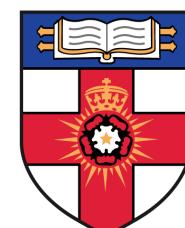
THINGS TO CONSIDER...

- Machine learning generated content needs framing / contextualisation.
- Challenges cultural heritage notions of “canonical” collection catalogue data.
- Important to approach critically.
- Specialist technical skills needed.
- Human expert input needed throughout.

THANKS

<https://www.sciencemuseumgroup.org.uk/project/heritage-connector/>

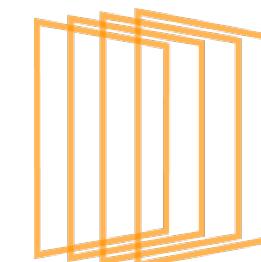
<https://thesciencemuseum.github.io/heritageconnector/>



SCHOOL OF
ADVANCED STUDY
UNIVERSITY
OF LONDON



Arts and
Humanities
Research Council



TOWARDS
A NATIONAL
COLLECTION

cogapp

