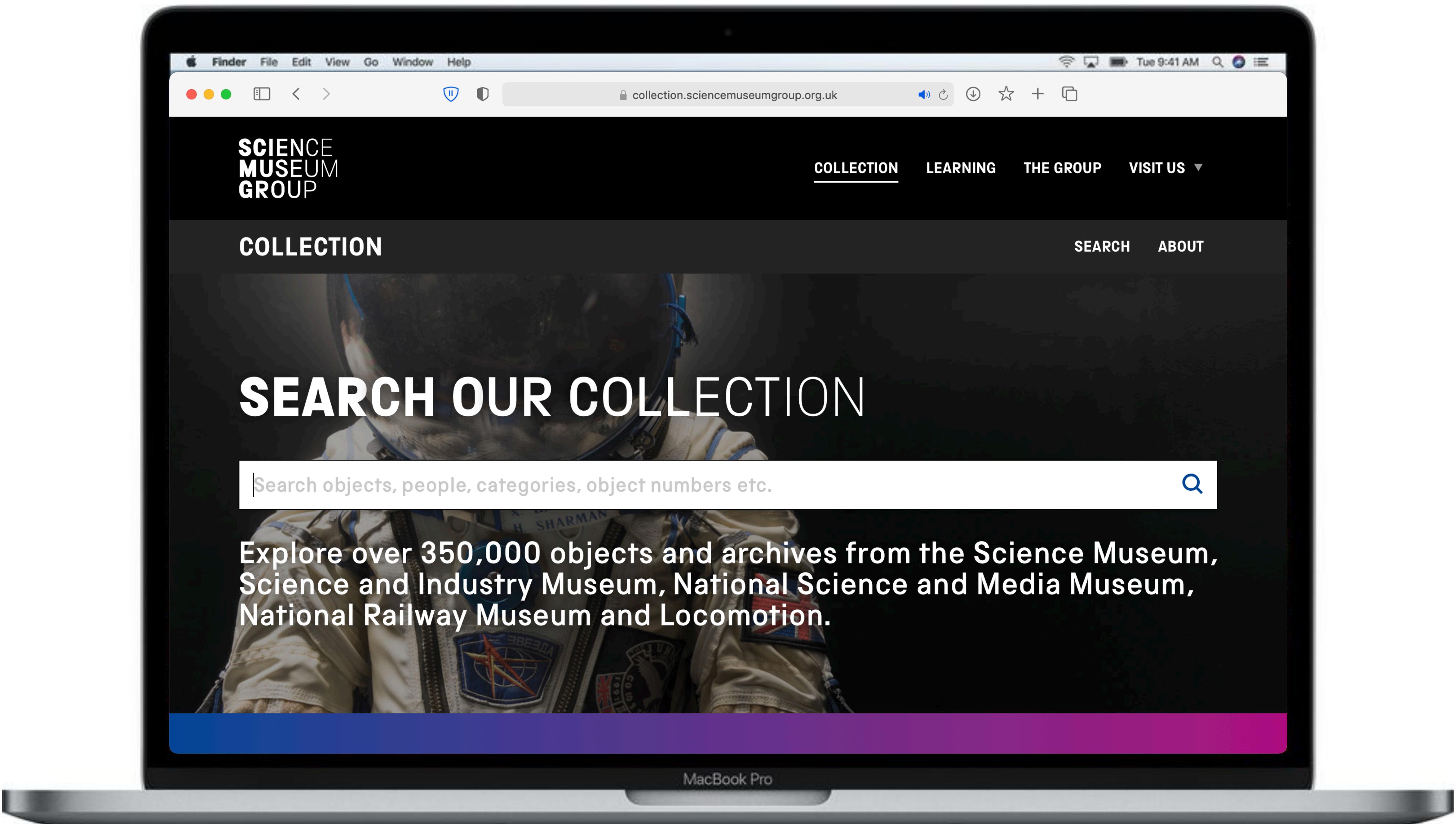


HERITAGE CONNECTOR: TRANSFORMING TEXT INTO DATA TO EXTRACT MEANING AND MAKE CONNECTIONS

Congruence Engine / Heritage Connector Event
John Stack, Digital Director
3 August 2023

SCIENCE
MUSEUM
GROUP



SCIENCE
MUSEUM
GROUP

COLLECTION LEARNING THE GROUP VISIT US ▾

COLLECTION

SEARCH ABOUT

SEARCH OUR COLLECTION

Search objects, people, categories, object numbers etc.



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.

MacBook Pro

Finder File Edit View Go Window Help

collection.science museumgroup.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

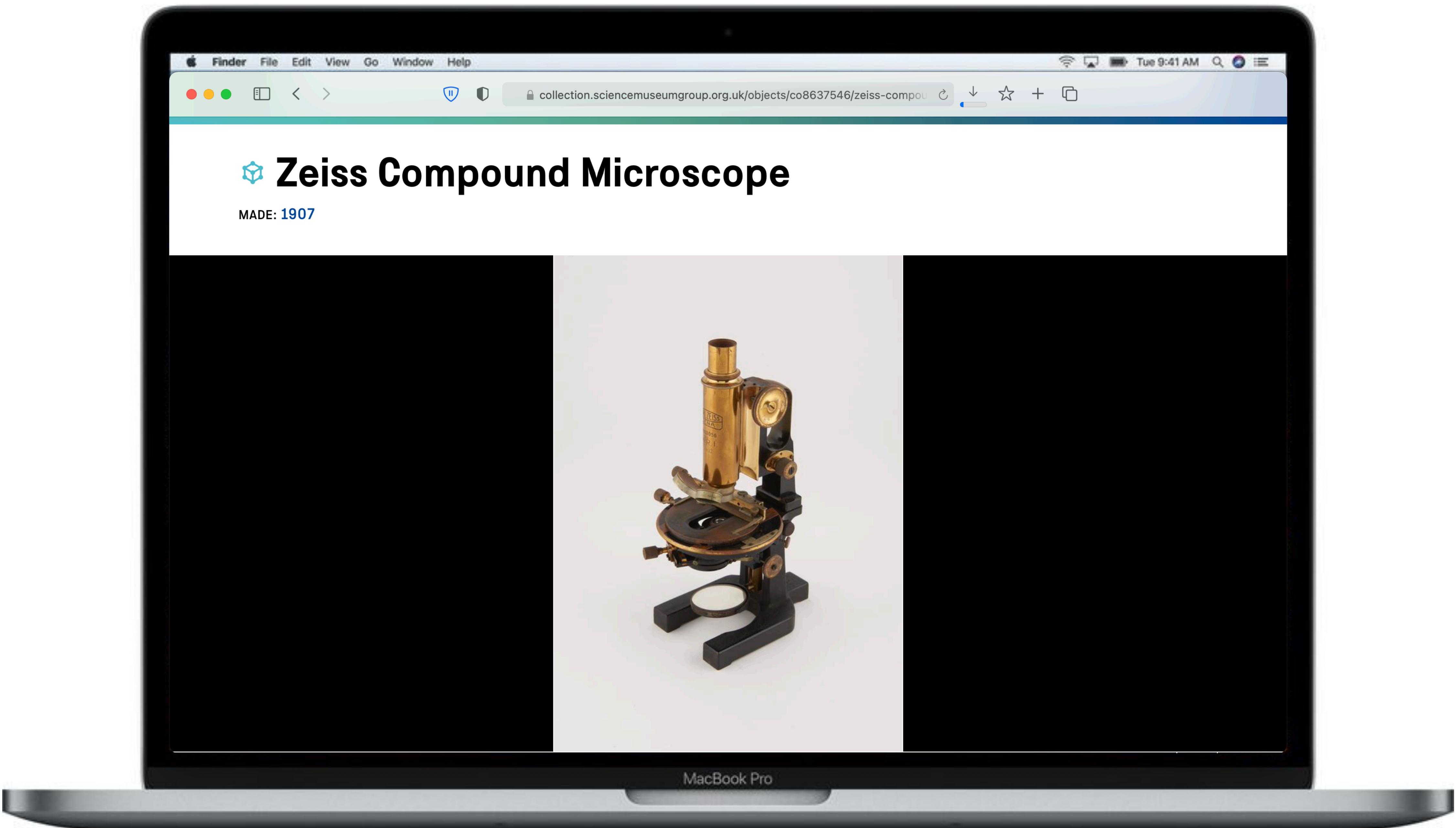
THE GALVANIC AND ELECTRO-MAGNETIC TELEGRAPHHS,
ON THE
GT. WESTERN RAILWAY.
May be seen in constant operation, daily, (Sundays excepted) from 9 A.M. to 5 P.M., at the TELEGRAPH OFFICE, LONDON TERMINUS, PADINGTON AND TELEGRAPH COTTAGE, SLOUGH STATION.
An Exhibition adapted by its numerous Visitors to be 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 islands of Australia, and nearly 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 persons in America, who will also then be enabled to speak or have a camera, in an incredibly short space of time, after the signal for his doing so 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 Tewksbury,)—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. Durand, the Agent for the Great Western Railway, Messengers in instant attendance to that communication received by Telephones, would

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



MacBook Pro

The image shows a MacBook Pro displaying a web browser window. The browser has a dark mode interface. The main content area shows the 'DETAILS' page for a Zeiss Compound microscope. The page includes fields for CATEGORY (Scientific Instruments & Research), OBJECT NUMBER (Y1991.49.2/1), TYPE (compound microscope), TAXONOMY (furnishing and equipment, tools & equipment, optical instrument, microscope), and CREDIT (Gift of Central Manchester Health Authority). To the right, a sidebar titled 'CITE THIS PAGE' contains a citation for 'Science Museum Group. Zeiss Compound'. Below it, the 'RIGHTS' section states that data in title, maker, and details fields are released under Creative Commons Zero, while descriptions and other text content are licensed under Creative Commons Attribution 4.0. It also links to 'Using our data'. The 'DOWNLOAD' section offers options to download the catalogue entry as JSON, view an IIIF manifest in an IIIF viewer, add the IIIF manifest to Animal Crossing Art Generator, and download the IIIF manifest directly. A note at the bottom of the sidebar states that records are constantly being enhanced and improved, but accuracy cannot be guaranteed.

DETAILS

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

CITE THIS PAGE

Science Museum Group. Zeiss Compound

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 Zero](#)

Descriptions and all other text content are licensed under a [Creative Commons Attribution 4.0 licence](#)

[Using our data](#)

DOWNLOAD

Download catalogue entry as [JSON](#)

View [IIIF manifest](#) in [IIIF viewer](#)

Add [IIIF](#) to [Animal Crossing Art Generator](#)

Download [IIIF manifest](#) [IIIF](#)

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 and discovery
- 5. Grasp the relevance of the content to the search
- 6. Navigate from that content to related content
- 1. Users rely on third-party search engines
- 2. Most collections opaque in terms of content
- 3. Limited content availability
- 4. Search relies on cataloguing and advanced search difficult
- 5. Content ranking is opaque and privileges certain content
- 6. Few or no links

LIMITATIONS

OPPORTUNITIES

LIMITATIONS

- Catalogue data is thin

OPPORTUNITIES

- Additional data can enrich collection records

LIMITATIONS

- Catalogue data is thin
- Human resources are limited, especially expert ones

OPPORTUNITIES

- Additional data can enrich collection records
- ML can add value (if we're satisfied with "good enough" and can define that)

LIMITATIONS

- Catalogue data is thin
- Human resources are limited, especially expert ones
- For users it's to get an overview of whole collection's content

OPPORTUNITIES

- Additional data can enrich collection records
- ML can add value (if we're satisfied with "good enough" and can define that)
- There's potential to present macro views of collections using knowledge graphs (KG)

LIMITATIONS

- Most records have few links to related content within and beyond collections

OPPORTUNITIES

- Potential to add valuable links at scale

LIMITATIONS

- Most records have few links to related content within and beyond collections
- Entry points of interest to users are not well handled in collections data

OPPORTUNITIES

- Potential to add valuable links at scale
- Potential to create new entry points for users (“Everything” related to X)

LIMITATIONS

- Although linked open data (LOD) has the potential to create new forms of research and discovery, they are onerous to create manually

OPPORTUNITIES

- Generate LOD via ML and explore potential to undertake such research, especially where previously impossible (too time consuming, too large scale to speculatively explore...)

PRIMARY RESEARCH QUESTION

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

collection.sciemuseumgroup.org.uk

SCIENCE MUSEUM GROUP

COLLECTION LEARNING THE GROUP VISIT US ▾

COLLECTION

SEARCH ABOUT

SEARCH OUR COLLECTION

Search objects, people, categories, object numbers etc.

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.

www.vam.ac.uk/collections?type=featured

BETA

HOME VISIT WHAT'S ON COLLECTIONS LEARN JOIN & SUPPORT SHOP

From the Collections

From ancient Chinese ceramics to Alexander McQueen evening dresses, take an incredible journey through 5000 years of human creativity with our online collections.

EXPLORE OUR COLLECTIONS

Search more than 1.2 million objects

Search by object, artist, maker...

SEARCH ▶

Add dates +

Not Secure — journal.sciencemuseum.ac.uk/browse/issue-16/se...

SCIENCE MUSEUM GROUP JOURNAL

BROWSE ALL ARTICLES

16 AUTUMN 2021 issue 16

View issue front page

Editorial by Sally MacDonald

Preserving skills and knowledge in heritage machinery operations by Pippa Cartt-Hornsby

A long engagement – railways, data and the information age by Robert Gwynne

The Whitworth: a place for industry and Art by Imogen Holmes-Roe

Reports and commands: deciphering a health exhibition using the SPEAKING mnemonic

Seismographs at Eskdalemuir Observatory, 1908–1925: tools for rethinking the origins of international cooperation in seismology

This article was written by Alexandra Rose
09-30-2021 | Cite as 10.15180/211607 | Research
[Seismographs at Eskdalemuir Observatory, 1908–1925: tools for rethinking the origins of international cooperation in seismology](#)
Published in Autumn 2021 Issue 16
Article DOI: <http://dx.doi.org/10.15180/211607>

blog.sciencemuseum.org.uk/ancient-greek-computing/

SCIENCE MUSEUM

BLOG HOME MUSEUM SITE

BY ROGER HIGHFIELD ON 15 DECEMBER 2021

ANCIENT GREEK COMPUTING

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



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	



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

↗ | > 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



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



> 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

> 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 Statio

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

Zero

WIKIDATA

Item Discussion Read View history Search Wikidata

Marks Tey railway station (Q2025932)

No description defined

In more languages

Statements

instance of railway station

image  Markstey1.jpg 648 × 429; 98 KB

named after Marks Tey

country United Kingdom

located in the administrative territorial entity Marks Tey

coordinate location  51°52'51"N, 0°46'58"E

operator Greater Anglia

THREE TECHNOLOGIES

1. Machine Learning (specifically Natural Language Processing, Entity Extraction)
2. Linked open data (LOD)
3. Knowledge graph

USER INTERFACES

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](#).

Mapping locations
Creating Google maps from places found in the Science Museum Group collection.

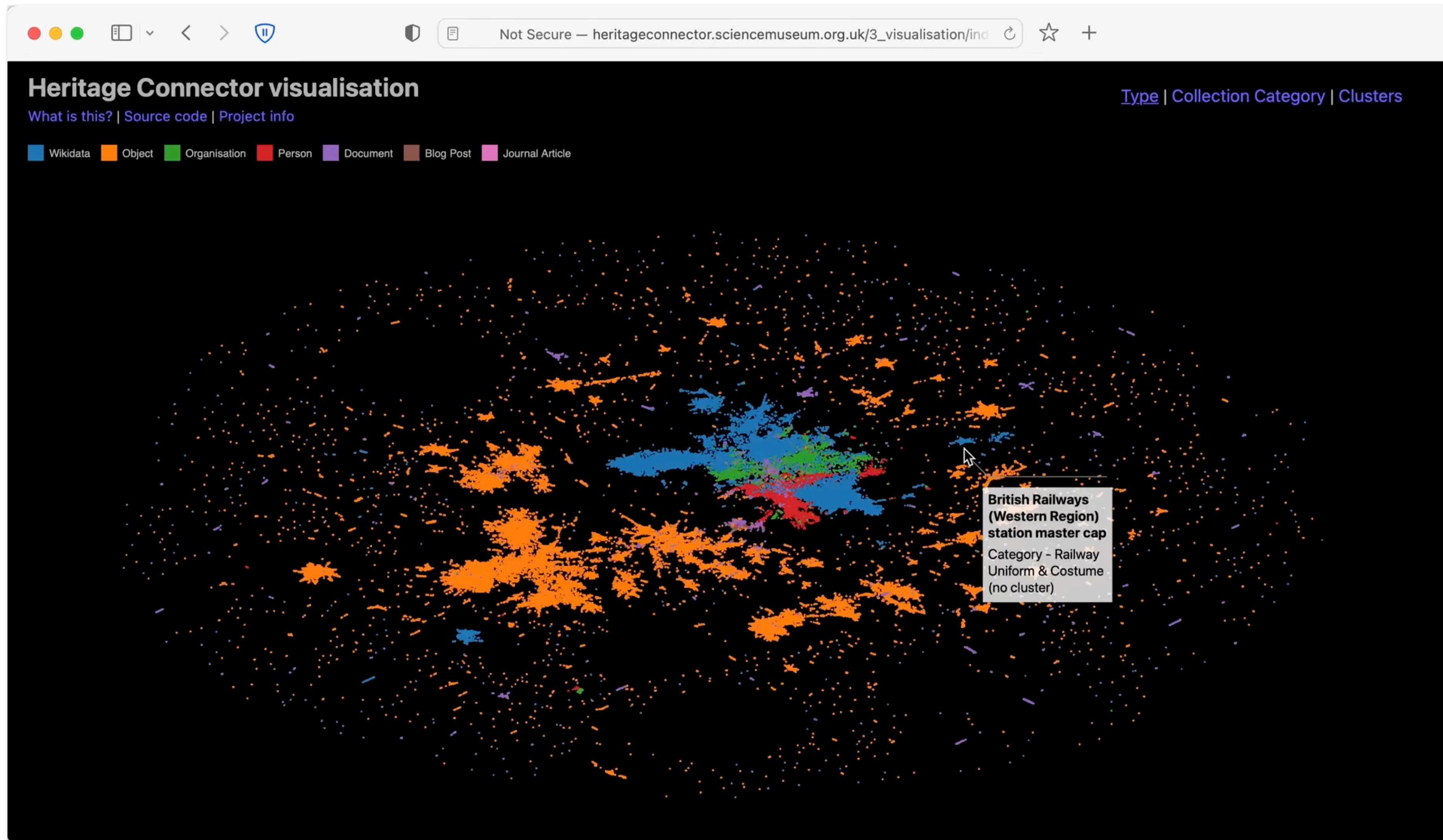
Metadata explorer
An interface for guided exploration of the Heritage Connector knowledge graph.

Visualisation
View a map of a knowledge graph containing the Science Museum Group's collection, blog and journal.

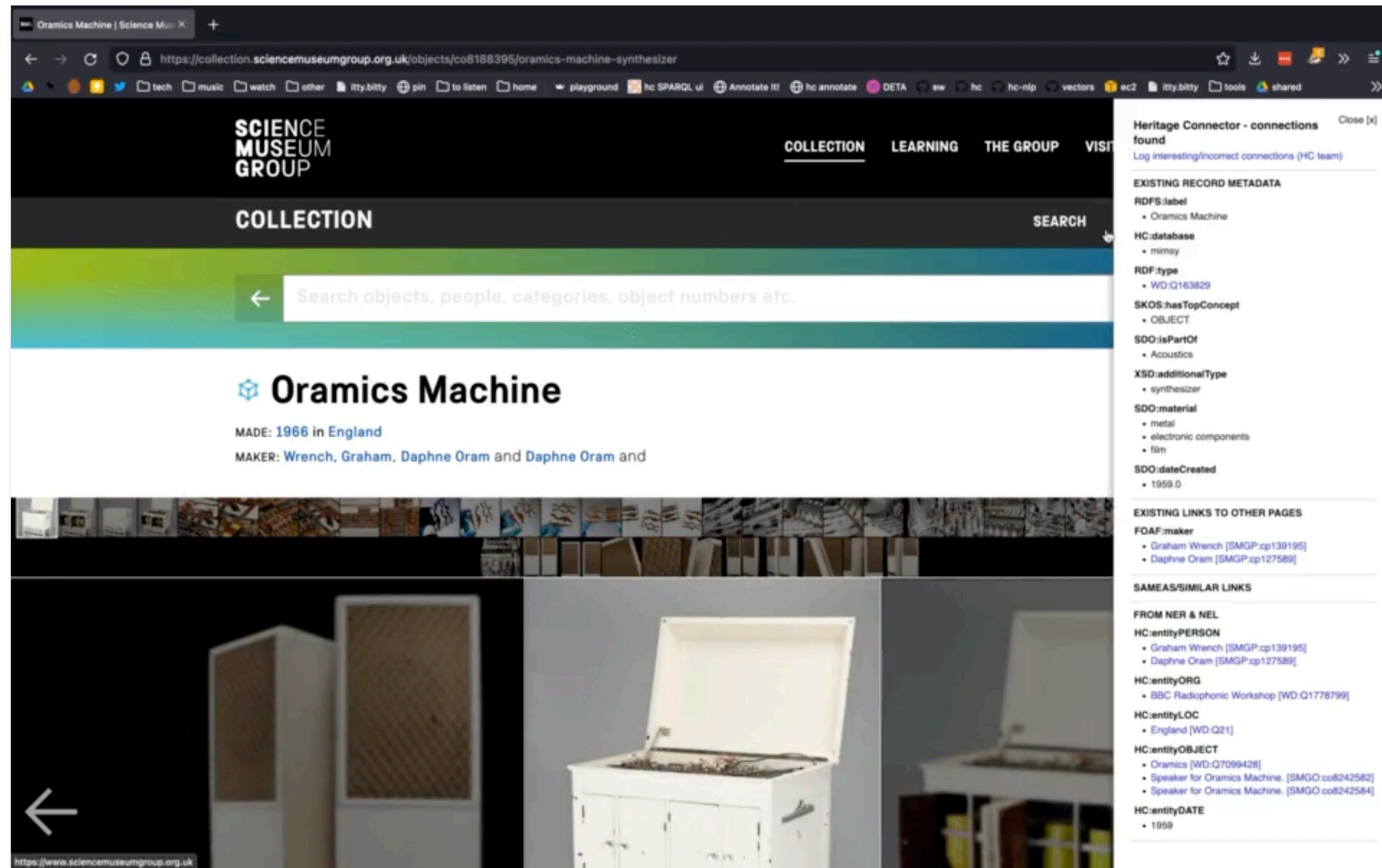
Visualisation - multiple collections
View a map of a knowledge graph containing the Science Museum Group's collection, blog and journal, and part of the V&A's collection.

Oramics Machine
A screenshot of a web browser showing a complex RDF query editor and a corresponding visualization of the data as a directed graph.

MACRO COLLECTION VISUALISATION



BOOKMARKLET



METADATA EXPLORER

Not Secure — 34.240.58.30:8010/view_connections

Heritage Connector Metadata Explorer

As part of the Heritage Connector project we've built a knowledge graph from the Science Museum Group and V&A collections using machine learning techniques.

This is an experimental interface designed to let you explore the connections in this knowledge graph, in a way that feels familiar.

[project website](#) | [blog](#) | [github](#)

Try some interesting starting points we've found...



Jimi Hendrix



Osaka



Indian Rebellion of 1857



Issey Miyake



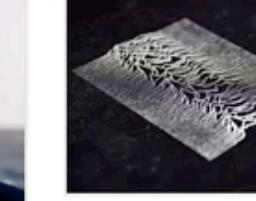
Tibetan alphabet



Brookhaven National Laboratory



John F. Kennedy



Joy Division



2012 Summer Olympics

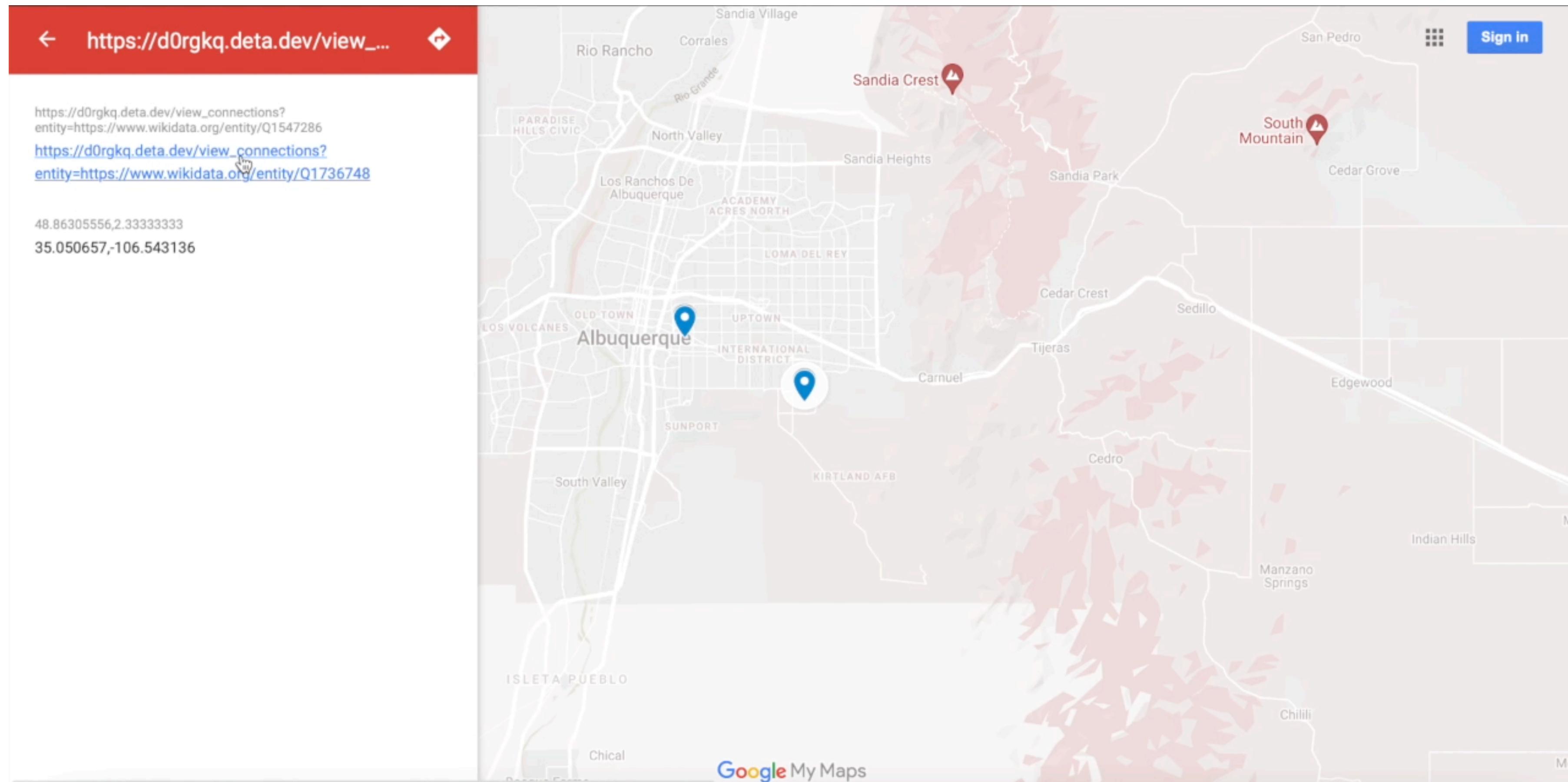


Queen Victoria

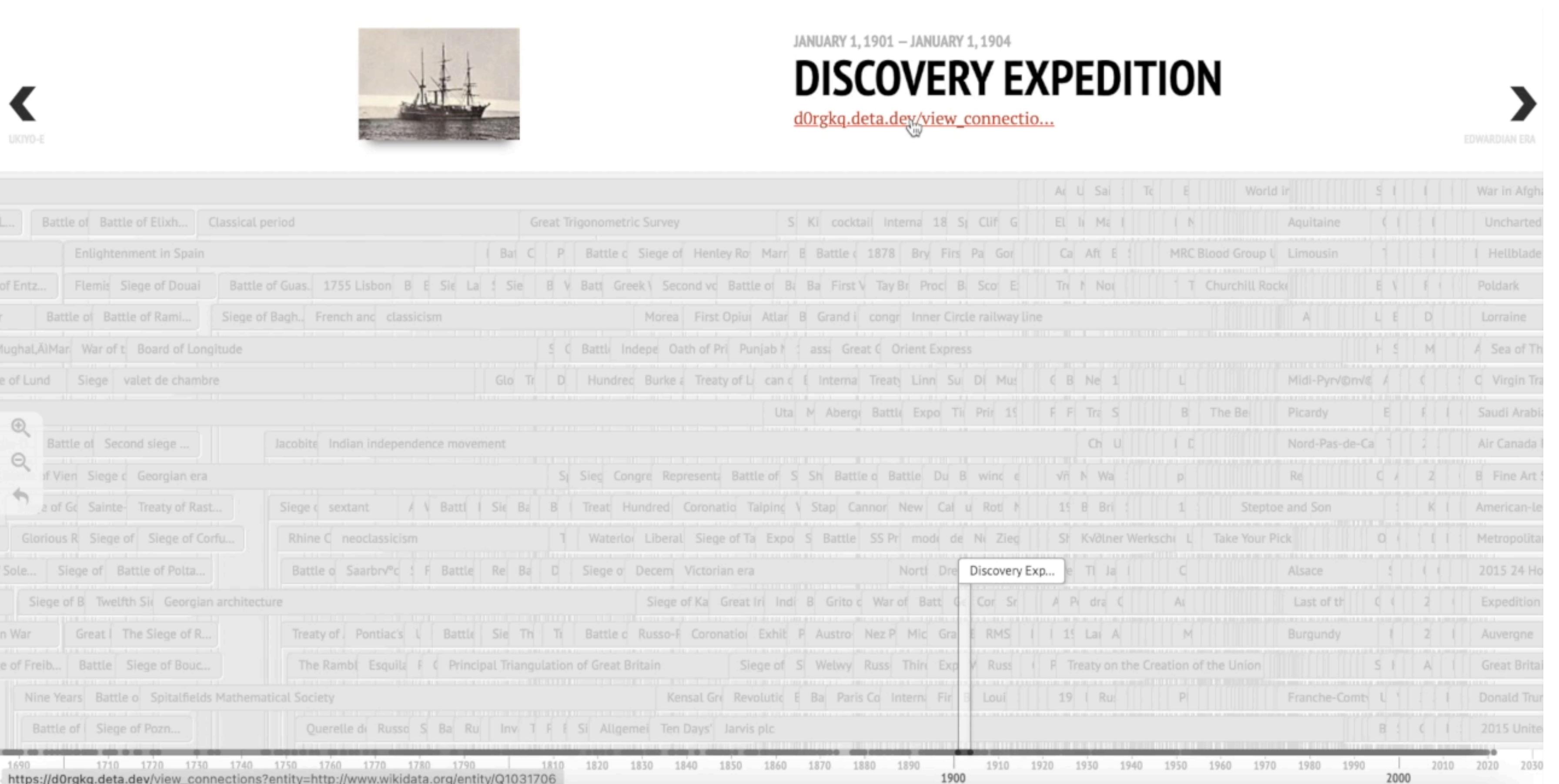
...or enter a URL

This URL can be from the Science Museum Group's [online collections](#), [blog](#) or [academic journal](#), or from some [select categories](#) in the V&A's online

MAP



TIMELINE



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 (with NER and EL).
- 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 ...

- Source data selection is important as data size grows rapidly.
- Data selection should be based on specific outcomes.
- Machine learning outputs need framing for users.
- Specialist technical skills needed.
- Challenges cultural heritage notions of “canonical” collection catalogue data.
- Human expert input needed throughout.

THANKS

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

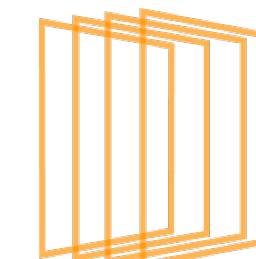
<https://thesciencemuseum.github.io/heritageconnector/>



SCHOOL OF
ADVANCED STUDY
UNIVERSITY
OF LONDON



Arts and
Humanities
Research Council



TOWARDS
A NATIONAL
COLLECTION

cogapp



ONE MORE THING ...

9 REFLECTIONS FROM 2023 ...

1. **User needs:** What value is sought and for whom? How to frame these questions?
2. **Wikidata:** Can only link to what's there. Gaps and biases. Analysis of strengths and weaknesses. Where to put effort to improve Wikidata for GLAMs?
3. **NER:** Strengths and weaknesses of entities extracted (people, organisations/companies, events, places, etc.). What other entities are valuable. How to extract these?

9 REFLECTIONS FROM 2023 ...

4. **Input data:** Selection needed and so tight definition needed. Which catalogue fields? Which collection areas? Consider extracting data from within objects (OCR, computer vision, speech to text, etc.)?
5. **Knowledge graph:** As storage mechanism in HC vs as end in itself? Further exploration of clustering? Identify gaps?
6. **Disambiguation:** Further exploration of this key element. Could crowdsourcing play a role (Reddit-style up/down noting by user)?

9 REFLECTIONS FROM 2023 ...

7. **Cataloguing:** Potential to input into cataloguing workflows as tool?
8. **Output data:** Collection is dynamic with overall glacial change and areas of rapid change. How to store, annotate and make useful output data (data likely has longer life than technology tools and methods)?
9. **User interfaces:** What user interfaces are most valuable and for whom? Do richer interfaces risk promising more than they can deliver?