

Resonating with different frequencies...

Thoughts on public humanities through
crowdsourcing in a ChatGPT world

Dr Mia Ridge, Digital Curator, British Library

@mia@hcommons.social @BL_DigiSchol@techhub.social

CAS 2023 Summer Symposium on Harry Shearer's Le Show Archives, August 2023

Overview

- My work in crowdsourcing / data science in GLAMs
- What can machines do?
- The *Le Show* archive (as described by Rosa)
- Why do we still need people listening to Le Show and other audio archives?

What is crowdsourcing / citizen science in GLAMs?

The basics: show people collection items and ask them to record information about them in some way

- Something in the image/sound, something they know, a relevant item they have
- Should contribute to a shared, significant goal or common good

(GLAMs – galleries, libraries, archives and museums aka cultural heritage organisations)

Heritage crowdsourcing as volunteering



...but convenient



<https://www.flickr.com/photos/jdevaunphotography/8456110245/> by Jason Devaun

Dora's lost data



"Hi, my name is Dora, and I'm a junior curator. It's my first day and I've made a big mistake – I accidentally deleted all the information we were going to add to our collections online. I need to re-label them, and quickly..."

Can you help? **Add words about the thing in the picture that would help someone find it on Google** – how it looks, what does, who might have used it – anything you can think of."

Scale model (1:10) of Swift gamma-ray burst satellite

Object from: Science Museum.

Date: c. 2000 Place: University of Leicester (Accession num: L2009-4034)



Add words ('tags') to describe this object

Tags

Tip: separate each tag with a comma, like this: tag, label, a phrase, name, names.

The Collective Wisdom Handbook

Perspectives on Crowdsourcing
in Cultural Heritage



Phil Ridge • Samantha Blakemore • Meghan Fenner

Wes Berglund Pryor • Ben Brunfield • Sara Brunfield • Sebosa Bunker • Jim Casey • Sonja Coleman
Sara Czubicka • Michael Haley Goldman • Kurt Luther • Austin Peat • Nick White • Brandon Wilkins • Pig Wilcox

<https://britishlibrary.pubpub.org>

Playbills *In the Spotlight*

Collection of over 230,000 printed sheets in 1,000 volumes

Minimal cataloguing: 'A collection of playbills from miscellaneous Plymouth theatres 1796-1882'

No information about individual playbills, performances, people, topics, places

Tasks ask for performance titles, genres and dates

Volunteers noticed much, much more, and shared information in comments



Designing for productivity and engagement

The screenshot shows a document viewer interface. At the top, a navigation bar contains 'MENU', 'IN THE SPOTLIGHT', 'Home', 'About', 'Take Part', 'Data', 'Browse Tags', and 'Discuss'. The 'Data' and 'Discuss' items are circled in red. Below the navigation bar, a toolbar contains several icons: a double-headed arrow, a question mark, an information icon, a download icon, and a share icon. This toolbar is also circled in red. The main content area displays a document page with text and a highlighted section. The highlighted section is a green box containing the text: 'PART II. Force of Hercules, With GROUND and LOFTY TUMBLING, By Master GARBUT, the first Performer in this Kingdom.' Below this, another section is visible: 'PART III. THAUMATACTICS; OR, THE WONDERFUL CONJURER!'. At the bottom left, there is a circular navigation pad with four directional arrows and a central button.

Transcribe the title

Type in the title exactly as displayed in the highlighted box on the page, including any subtitles and punctuation. You don't need to include any genre or description.

TITLE

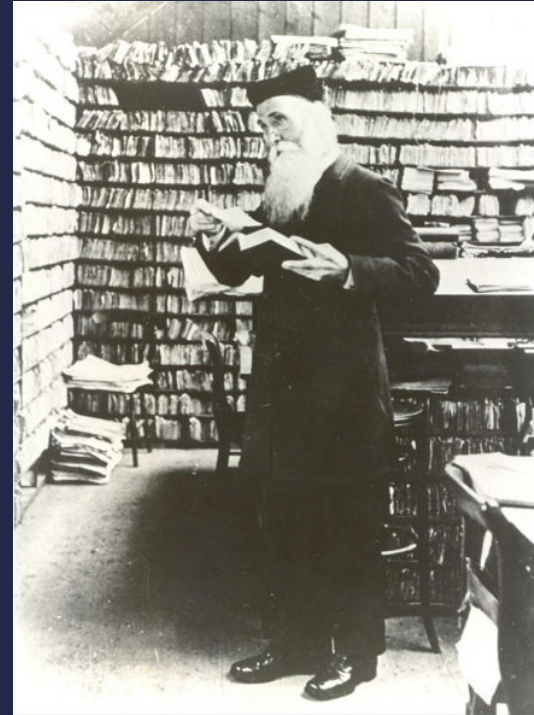
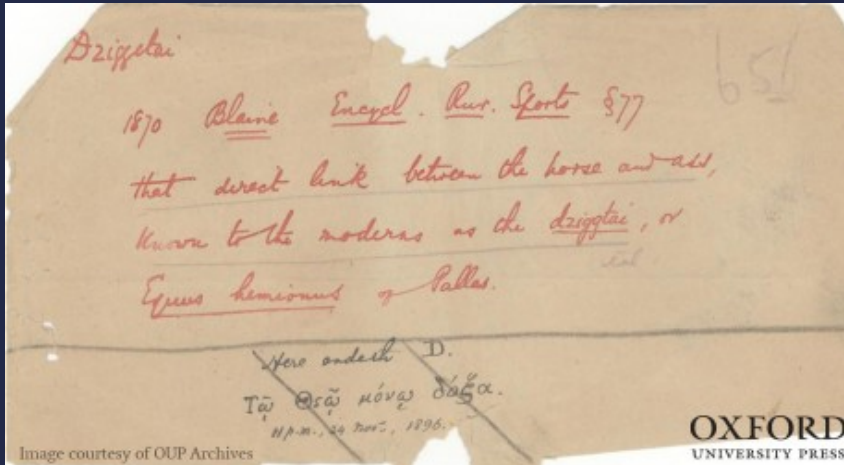
[Tag this sheet](#)

[Something's wrong](#)

[Seen something interesting?](#)
[Add a note](#)

Save and Continue

Technology helps, but
ultimately
crowdsourcing is all
about people

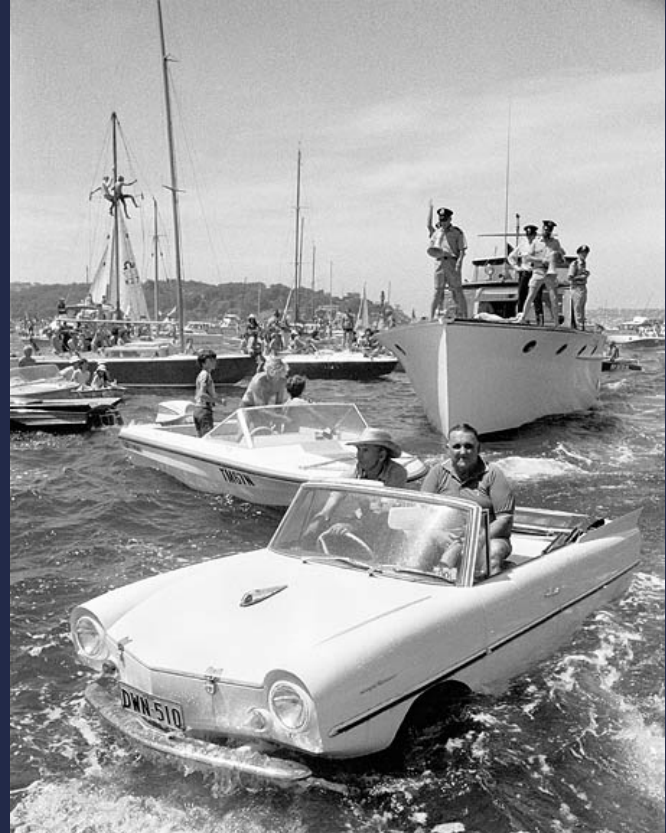


James Murray, editor, OED, with
contributor slips

My favo(u)rite tip: design for motivations

People crave:

- satisfying work to do
 - the experience of being good at something
 - time spent with people we like
 - the chance to be a part of something bigger
- Jane McGonigal, 2009



Crowdsourcing brings collections to life

Close, engaged attention to 'obscure' collection items

Opportunities for lifelong learning; historical and scientific literacy

Gathers diverse perspectives, knowledge

Living with Machines: responding to the growth of digital scholarship and data science



An opportunity to:

- Collaborate with The Alan Turing Institute, the national institute for data science and artificial intelligence, based in our St Pancras building
- Collaborate with subject and methodological experts, building on the Library's expertise in research services and public engagement
- Understand the potential and challenges of AI / machine learning for cultural organisations
- Build on digitisation work

Some methods from *Living with Machines*

- Analysing text with Natural Language Processing (NLP) and Large Language Models (LLMs):
 - Toponym resolution – finding and disambiguating place names in text
 - Nominal linkage - algorithmic record matching to find individuals between census years
 - Linguistic research – detecting when machines are assigned agency; semantic shifts in meaning across time, place
 - Contextualising data with metadata (looking for bias e.g. liberal/conservative)
- Training machine learning to analyse visual features on maps
- **Embedding public engagement through crowdsourcing** – finding articles about accidents involving machinery and annotating details about them; defining 'machines' through transcription and classification; learning from conversations influenced our exhibition
- Small scale 'human in the loop' experiments (labelling images as 'ad or not' to create training data for machine learning to distinguish adverts from news stories)

Thank you everyone who helped complete our projects!

FINISHED!

ST. HELENS BICYCLE CLUB.—On Thursday a handicap race between members of the St. Helens Bicycle Club took place. The course was from the Bird-i'-th'-hand to the King's Arms, Prescott, thence to Rainhill Bridge, returning to the Bird-i'-th'-hand *via* Nutgrove, the distance being six miles and a half. Eight prizes were offered, the first being a handsome silver cup, presented by Mr. E. Bowen, and the remainder given by the club. Mr. E. Bowen acted as judge, and Mr. W. J. Ashton (president), Mr. Fred. R. Nuttall (vice-president), Mr. S. Wolfenden (captain), and Mr. C. Rowe (hon. sec.) as starters. At eight o'clock, the time announced for the start, thousands of people from St. Helens assembled near the Bird-i'-th'-hand, and there were many present on horseback and in vehicles. The weather was fortunately quite favourable, which made up for the disappointment occasioned on the previous day, when the race was originally announced to take place. The following members of the club started: J. Haslam, 7 min. allowed; E. Haslam, J. Deane, and F. Taylor, 5 min. 3 sec.; W. Swift,



TASK

TUTORIAL

We searched for old articles that mention 'machine' and 'accident' keywords. Help us check the results.

Does the image mention a specific accident involving machinery?

Check Help for tips.

There is only one article on the image and it mentions a specific accident involving machinery

There is more than one article on this image and at least one mentions a specific accident involving machinery

There are no relevant articles on this image

The article text is too hard to read

NEED SOME HELP WITH THIS TASK?

Done & Talk

Done





Thank you everyone who helped complete our projects!

Living with Machines Talk

[Living with Machines Talk](#) > General project discussion

General project discussion

Puzzled by an instruction? Got a question or an idea for improving things? Share here!

Please [sign in](#) to create discussions

[Our current project: how did machines change accidents?](#) 5 Participants
 Polarpat22 *2 months ago* That sounds like fun! I'm looking forward to hearing how you get on. 78 Comments

[Has this project changed how you think about machines in the 1800s?](#) 5 Participants
 yladly *2 months ago* There have always been people who are sure they can risk anything, despite wa... 16 Comments

[What 'machine' words changed when?](#) 6 Participants
 miaridge RESEARCHER TEAM *4 months ago* We've decided that 'train' isn't likely to yield interesting results in looki... 51 Comments

[Postscript about railway fatalities](#) 2 Participants
 Nomad Purple *2 months ago* I think the difference is that most of those are deliberate, only 23 are unin... 2 Comments

Recent Comments

Popular Tags:

- [fatalaccident](#)
- [accident](#)
- [railway](#)
- [rail](#)
- [grisly](#)
- [fire](#)
- [boiler](#)

**Given developments in 'AI'...
(machine learning)**

What can AI/data science do for audio?

- Transcribe speech for text-based search, methods
- Detect some concepts, entities, emotions -> metadata for findability
- Support 'distant reading'
 - Shifts, motifs, patterns over time
 - Collapse hours, years – take time out of the equation
- Machine listening?
 - Use 'similarity' to find sonic (not text) matches?

WORLD SERVICE RADIO ARCHIVE PROTOTYPE

Search e.g. from our own correspondent

About | Feedback 45

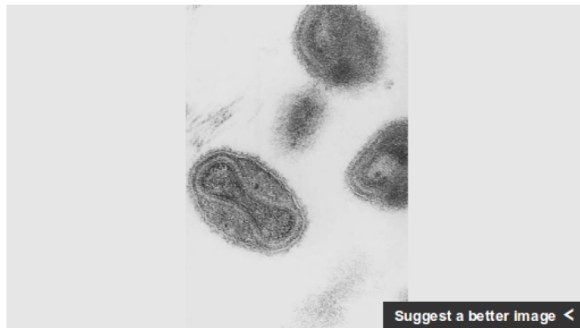
MON 16 MAY 2011 | SCIENCE AND HEALTH

Discovery
Smallpox

Should we deliberately make a living organism extinct? What if it's the Smallpox virus, that killed around 300 million people in the 20th Century. The cases for and against destroying the remaining stocks, from Edward Hammond & Raymond Weinstein

🕒 17 minutes

⚠️ FLAG THIS EPISODE AS INAPPROPRIATE

[Add to homepage](#)[Suggest a better image <](#)

00:00/17:28

Tags

These tags describe this programme so that other listeners can find it. You can help us improve these tags by voting down the ones that do not apply to this programme and voting up those which best describe this programme.

✎ Add tag

TAG	RATINGS	SOURCE
Smallpox	15 0	Synopsis and audio
Infectious disease	12 0	Audio
Eradication	8 0	User
Virus	8 0	Audio
Vaccine	8 1	Audio
Edward Jenner	7 0	User

Le Show (as described by Rosa)

- A massive 'portal' of 'conceptual and sonic hyperlinks to late-20th- and early-21st-century news and culture'
A 'polyphonic cornucopia of words and characters, lyrics and arguments, fact and folly'
- 'resistant to datafication'
- With koine topoi – issues of common or public concern

The screenshot displays the Harry Shearer website interface. At the top, there is a navigation bar with a 'Listen' button featuring a play icon. The main content area shows a date 'January 24, 2021' and a link to 'Download Full Episode'. Below this, a featured segment titled 'News of Bees: Bees aren't getting enough sleep' is shown with a play icon and a progress bar. A 'MUSIC & SEGMENTS' section lists various audio clips with their start times and titles, such as 'Open/ Tribute to Larry King' and 'Come Self-Isolate With Me' by Harry Shearer. At the bottom, a 'PLAYLIST' section lists several tracks. The website's branding 'HarryShearer.com' and navigation links for 'Le Show', 'Store', 'Projects', and 'Media' are also visible.

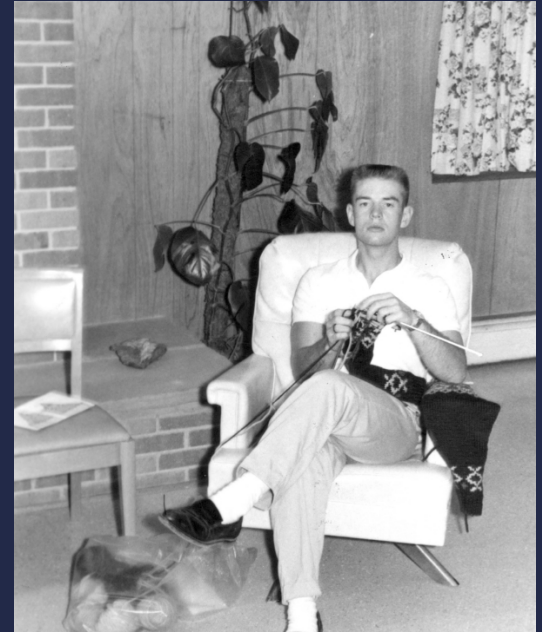
Why let machines have all the fun?

- People can hear a richer range of emotions, topics and references, recognise impersonations and characters
-> better metadata, findability

Why let machines have all the fun?

Could access to history in the raw, 'koine topoi' be a super-power?

Individual learning via crowdsourcing contributes to an informed, literate society



Jay Yoder with His Knitting

Thanks for listening. Questions?

Dr Mia Ridge

Digital Curator, British Library

@mia@hcommons.social

<https://bl.uk/digital>

<https://livingwithmachines.ac.uk>

<https://britishlibrary.pubpub.org/>

