TOWARDS A NATIONAL COLLECTION

UKRI Arts and Humanities Research Council

FIRST REPORT
DISCOVERY PROJECTS

The Congruence Engine:
Digital Tools for New Collection-Based Industrial Histories

Science Museum Group | University of Leeds
University of London | The Turing Institute | Historic England
University College London | University of Liverpool
British Film Institute | National Museums of Scotland

OCTOBER 2022

Table of Content

Authors

Tim Boon (Science Museum Group), Anna-Maria Sichani (University of London)

Executive Summary

Nine months into our three-year programme, *Congruence Engine* is on time and on budget, addressing the stated aims and objectives; the project team is working selectively with the project partners (in the expectation that all will be enrolled across the duration of the project). Our systemic action research methodology is working well to organise the investigation and, even now, tentative findings – which will be the subject of further exploration – are emerging. This year, we have two main critical paths in the investigation: the textiles research strand (initially defined as a 'pilot') and the creation of a special issue of the *Science Museum Group Journal* that will articulate the scope of, and aspirations for, the whole project. Our project blog is up and running, with eight posts so far, and we are using BaseCamp as our project management tool, accessible to the extended team.

The 'personality' of *Congruence Engine* within the set of five Discovery Projects is emerging: it is the project that is exploring and experimentally enacting what it would mean to potential future users of a national collection to be able to do historical and curatorial work in a world of cross-collection accessibility. This has entailed a modest change of technical focus, away from using AI to construct a single knowledge graph (which we already achieved in our *Heritage Connector* Foundation Project) towards more experimenting with off-the-peg digital humanities programs and techniques. This is closely linked to a deepening understanding of the 'material conditions' of the project, which include the differing time commitments of investigators and researchers within the project, the distribution of team members across the country and the consequent more extensive than expected use of online and hybrid methods, navigating the inherent strengths and weaknesses of the different means of communication open to us.

Staffing: The project is currently 83% fully staffed and working well. The project and team have adjusted well to some unexpected difficulties of staffing:

- The impossibility of recruiting a full-time three-year data scientist has led to the pooling of the wide range of digital expertise to support the investigation and the recruitment of Alex Butterworth as Senior Research Fellow, Digital Public History at 30%. As the investigation is showing, a data scientist was not needed at the start of the project, and initial enquiries show we will be able to bring in data science expertise as specific smaller work packages from year two.
- The tragic early death of our Energy History Research Fellow Cameron Tailford in August has been hard for the team, and requires a new recruitment.
- The resignation of our Project Manager after 5 months requires a new recruitment, which is under way. In the meantime the core duties of this post are being ably delivered by Carol Chang, SMG Research Support Officer, on secondment.

Progress: With setup achieved, the Opening Conference, 9-11 February in Saltaire and Bradford, convened the wide research team (investigators, researchers and staff from Collaborating Organisations and Project Partners) in a discussion of opening expectations and project possibilities. The event was conducted under the action research methodology, and it led to commencing work on the project's first industrial theme, textiles, in the form of a 'pilot' gathering of data within the Omeka-S platform. With this in place, June 20-22 we held an engagement workshop in Leeds, bringing together a subset of the research team with community

partners, members of the Saltaire History Collection. This led to six mini-investigations led by team members in close discussion with our community partners, using different digital techniques to link machines, pictures, oral history and online demographic data as exemplar projects of how to link the national collection as it touches on the history of the woollen trade in Saltaire, Manningham and Bradford more widely. A reflection and planning workshop was held in London on 27th July to hear reports of the mini-investigations and to work through their implications for the next stage of the research. We are now planning in detail for the second round of textiles work, which will extend to cotton production. In parallel, we are writing a special *Congruence Engine* issue of the *Science Museum Group Journal* to be published online in December. This is a core line of the research, deliberately articulating early in the project how the participants see the potential and practicalities of linking collections and their aspirations for what the project can do. We expect to publish up to 14 contributions, which – reflecting the concerns and research style of the project – will vary in format; contributions include conversations, picture essays, a film and opinion pieces and reviews alongside more conventional single-author academic articles.

Research Results: Even at this stage, there are emerging conclusions that we will test as the project develops. These include:

- Coming to consider the 'national collection' more as something that is brought into existence by acts of linking than as an entity that exists already waiting to be revealed; we have been speaking of this as using 'to national collection' as a verb.
- Realising that people wanting to do history with collections are very often working at the small, 'hand-stitched,' level. The project therefore has to display vanguardism in showing ways in which working digitally at scale can enable historical and curatorial research. The same applies to linking collections of different *kinds*, including objects, archives, films and broadcasts in an intermedial fashion: *Congruence Engine* provides the opportunity to demonstrate what could be done.
- Linking collections – of objects, pictures, etc, whose catalogue descriptions are often thin and which rarely record histories of everyday use – requires the 'connective tissue' of bigger historical data – such as street directories, the Census, etc – to do the kinds of social 'history from below' that many community historians would like to pursue. Such 'big data' sources should be seen as parts of the national collection we are to connect.
- The acts of creating a digital national collection require the effective interdisciplinary collaboration of several differing work cultures – digital, humanities, curatorial, participative action-research. It is easy for misunderstandings to arise if these differences are not appreciated. We are beginning to address this via a small collaboration with social scientists.

Abstract

The capacity to make strong connections between historical objects and sources lies at the heart of this project as it does in the everyday museum and historical practices that it is designed to support. Curators creating displays combine artefacts, images, audio-visual materials and histories. Family and local historians connect records of ancestors and localities to establish their genealogy or to understand the past of where they live. Academic historians patiently and critically connect a diverse range of archive sources with existing literature to tell new stories about the past. All rely on connecting different fragments of the past as they weave the tapestries of narrative that constitute our local and national histories. *The Congruence Engine* will create the prototype of a digital toolbox for everyone fascinated by the past to connect an unprecedented range of items from the nation's collection to tell the stories about our industrial past that they want to tell. This project explicitly works with collections that are generally represented by weak data. In place of the two-dimensional ranked list of search engines, we aim, with 'The Congruence Engine', to model a world in which users will be able to explore data neighbourhoods where a great diversity of information about heritage items that are deeply relevant to their investigations will be readily to hand – museum objects, archive documents, pictures, films, buildings, and the records of previous investigations and relevant activity.

Aims and Objectives

1. Via an action research methodology to use real-world historical enquiries of community-, museum- and university-based historians and curators to hone digital tools needed for proof of concept of a linked national collection, using the example of industrial history.

This is well under way; the methodology was applied at the opening conference in February, and has been pursued in the first round of the textiles investigation. We have, so far, been through one Action Research cycle of plan-act-observe-reflect and are embarked on the second cycle. One adjustment: in the methodology's spirit of responding to emergence, we have made the digital contribution symmetrical with that of the curatorial/research 'side,' rather than the implied specifier-supplier relationship implied in the original formulation.

2. To conduct three industrial sector-based investigations using heterogeneous collections.

    - Textiles, energy and communications are here used to map the changing history of industry, society and culture over 250 years.

    The textiles investigation is halfway through; in the autumn we will extend from wool in Yorkshire to cotton and other fibres in Manchester and Lancashire. Co-Investigators for the energy and communications strands are well advanced in developing frameworks for their areas too, and are also considering cross-project themes linking two or three together (with final outputs in mind), such as the changing use of energy in the textile trade for example.

3. To represent museum culture and STEM history within the set of Discovery Projects.
    The differences between museum styles of research ('museum as method') and those of the Academy have risen up the project's agenda, as kinds of practice that draw on many collections and on differing media are typical of museum work (especially in display), but less so in the universities. Histories of science, technology, engineering and medicine will also occur in the other discovery projects to a lesser extent, but here are the core concern. Furthermore, the collections we can link are products of the collecting histories of institutions; in the case of technical museums, understanding how they have conceptualised technology in relation to science is key to unlocking understanding of the collections that exist.

4. To pursue kinds of interdisciplinary and intermedial research that are specifically enabled by the uniting of heterogeneous kinds of heritage items and their data.
    The research is necessarily interdisciplinary, even at the level of combining humanities with digital expertise; the action research adds a third kind of practice to the mix, whilst recognising that museum modes of research (see 3) differ from those of universities adds a fourth. We are committed to making the research inter- not just multi-disciplinary and we have initiated some work with social scientists to explore this (see 9 d.). Intermedial research styles, currently being articulated in submissions to the Science Museum Group Journal special issue, will be pursued from the next stage of the project (the second half of the textiles investigation).

5. To conduct the investigation UK-wide : Bradford, Manchester, Newcastle, London, Edinburgh, etc.
    In addition to the sites of planned activity, we have Project Partners from across the UK, including National Museums Northern Ireland and Amgueddfa Cymru.

6. To create appropriate outputs for all audiences: general and enthusiast public and scholars.
   Interested individuals can already read the project blog (eight issues written as of 17th August). The special issue of the Science Museum Group Journal *will be available open access from the end of 2022 to colleagues anyone who wants to access it. We have begun planning for our interactive exhibit to be launched at the Science and Industry Museum in Spring 2023.*

7. To create, enable and sustain a multidisciplinary team with appropriate skills to deliver the project aims.
   The Action Research methodology is working well to convene the team. Still working hybrid, we are moving to maximise the opportunities for working together in person, for example by fixing a team day (Wednesday) for those members of the team readily able to travel to London, and by opportunistically holding meetings around project events. We have a digital skills working group, poised to deliver training this autumn. Some training (for example in Omeka-S and OpenRefine) has already been undertaken.
8. To enhance the cross-disciplinary skills of participants, especially the employability of early career researchers (ECRs).
   The ECRs (and indeed the whole team) are working across disciplines, learning digital, curatorial, historical and participative skills. All team members are able to draw on each other's wide and deep experience.

9. To apply digital – including AI – techniques to thin and inconsistent collections data to create research-valuable webs of connection between collections, whilst addressing biases in collections, data and AI techniques.
   Even at this early stage, we have used a wide variety of digital tools including, particularly: Omeka, Easy Refine, Zotero, GitHub, BaseCamp, SpaCy (Named Entity Recognition), Gephi (network analysis of census families), Qgis, Kepler.gl (spatial analysis) and drew on the Campop data set for work on migration. The PyTesseract code has been used to OCR text. We have experimented with the Neo4j graph database with its inbuilt graph visualisation feature. Also, bespoke code has been applied to (a) extract n-grams (uni-, tri-, bi-) for initial probe of free text content (eg collections data field) and to start building vocabularies; (b) reconstruct genealogies from complex Census household-relation descriptions; (c) scrape website of hierarchical occupations descriptions into cleaned and tabulated form. We plan to use Transkribus to test handwritten text transcription once we have identified suitable material.
   More generally, we are repeatedly discussing the characteristics of 'hand-stitched' historical work and looking for opportunities to move up the scale to bigger data and the application of machine learning techniques. Here the intention is to commission some work within the next six months; we are keeping a short list of likely research areas, including automatic parsing and OCRing trade- and street- directories.

10. To nurture dialogue in the digital humanities space between heritage/ humanities and data/computation.
    See 4 above.

11. To work with other Discovery projects and AHRC to maximise the benefits and legacy of TaNC.
    Three of the other Discovery Projects have agreed to take part in refereeing the special issue of the Science Museum Group Journal. We look forward to further collaborations as the project mature.

Partnership structure

Funded Partners

- Science Museum Group, base for Principal Investigator (PI) Tim Boon and Co-Investigators (Co-is): John Stack, Jamie Unwin and Dave Patten
- University College London, base for Co-I Jon Agar (Communications history)
- University of Leeds, base for Co-I: Helen Graham and Research Fellow Arran Rees (Action Research), and Co-Is Simon Popple (Digital Community engagement) and Graeme Gooday (Energy History),
- Liverpool University, base for Co-I William Ashworth (Textiles history)
- School of Advanced Studies, University of London, base for Co-I Jane Winters and Research Fellow Anna-Maria Sichani (Digital Humanities)
- Historic England, base for Co-I Wayne Cockroft (Historic Environment)
- British Film Institute, base for Co-I Patrick Russell (Film)
- National Museums Scotland, base for Co-I Geoff Belknap (Communications history)
- Bradford Industrial Museum, Collaborating Organisation, base for Vicky Shaw (Textiles history)
- Discovery Museum, Collaborating Organisation, base for Kylea Little (Energy history)
- Madlab, base for Asa Calow (Digital technique and participation)
- Wikimedia UK, base for Stuart Prior and Daria Cybulska (interactions and collaborations over Wikidata)

Unfunded Project Partners (data providers and collaborators)

- Grace's Guide to Industrial History
- BBC History and Heritage / Programme Index
- The National Archives
- Saltaire World Heritage Education Association
- BT Archives
- Birmingham Museums Trust
- Tools of Knowledge Project
- National Museums of Northern Ireland
- National Museum Wales
- The National Trust
- Society for the History of Technology Bibliography
- History of Science Society: Isis Bibliography
- Victoria and Albert Museum

Staffing structure

Each of the eight work packages is the responsibility of an investigator (some with other investigators), and most have an associated research fellow. There have been three challenges in the staffing at this point: Tragically, the young postdoctoral research fellow working on the energy strand died unexpectedly in August, and we shall now have to recruit to fill his post. Earlier, it proved impossible to recruit a full-time data scientist for the whole project (we are now looking to let smaller contracts for specific bodies of work), and we are also having to re-fill the Project Manager post following the resignation of the first incumbent (duties are currently being undertaken by secondment of SMG Research Support Officer). Below are the top line research question for each (some edited reflecting development within the project):

Umbrella WP1: Tim Boon, PI; Project Manager (vacant), Project Coordinator: Nina Webb-Bourne

- What could a national collection of UK industrial history created with the application of Digital Humanities and AI tools, and drawing on repositories of all kinds and scales, look like?

Collections Data and AI WP2: John Stack (oversight), Jamie Unwin, Senior Research Fellow Alex Butterworth, Asa Calow (MadLab), and machine learning work packages to be defined and contracts let

- How can we apply existing digital humanities tools and AI techniques to support real-world historical enquiry?

Digital Humanities WP3: Jane Winters; Research Fellow, Anna-Maria Sichani

- How do technology choices affect research processes and introduce biases, with what ethical implications? How do visualisation and digital mediation / remediation affect historical practice, including historians' qualitative and quantitative readings?

Participatory Action Research WP4: Helen Graham; Research Fellow, Arran Rees

- How can systemic action research technique be used to understand the challenges of realising a national collection of industrial history?

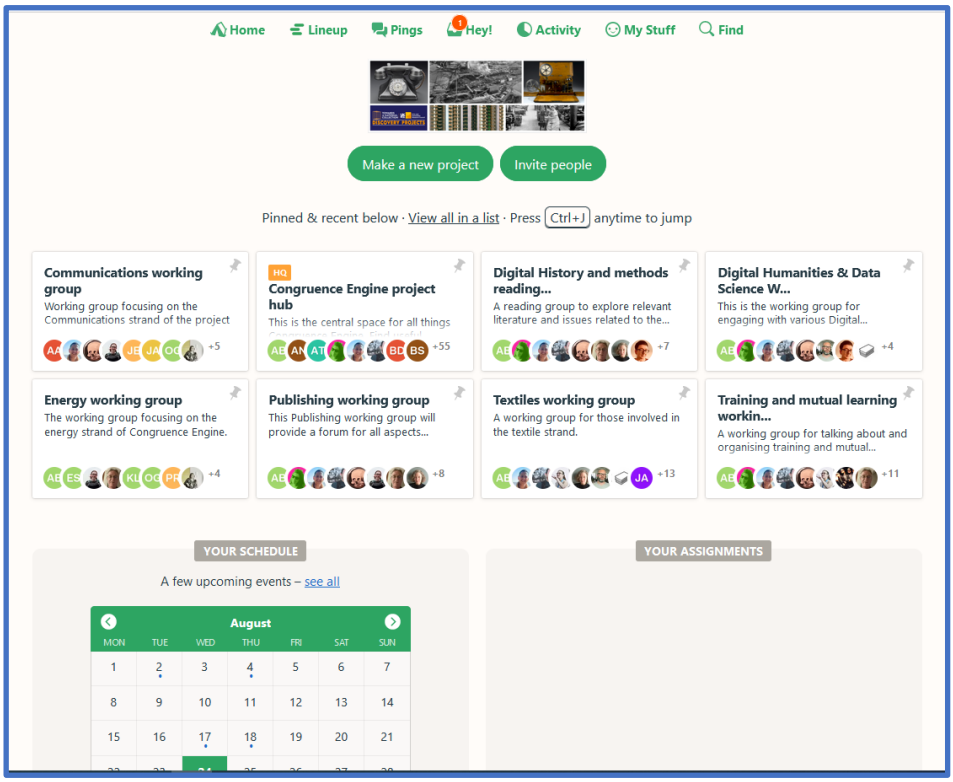
Digital exhibit WP5: Dave Patten

- How can we best use multiuser interactive installations with projected displays in museum contexts to present data and narratives in a way that enables visitors to interact, and the project to gain evaluated insights into, the effectiveness of digital tools with users?

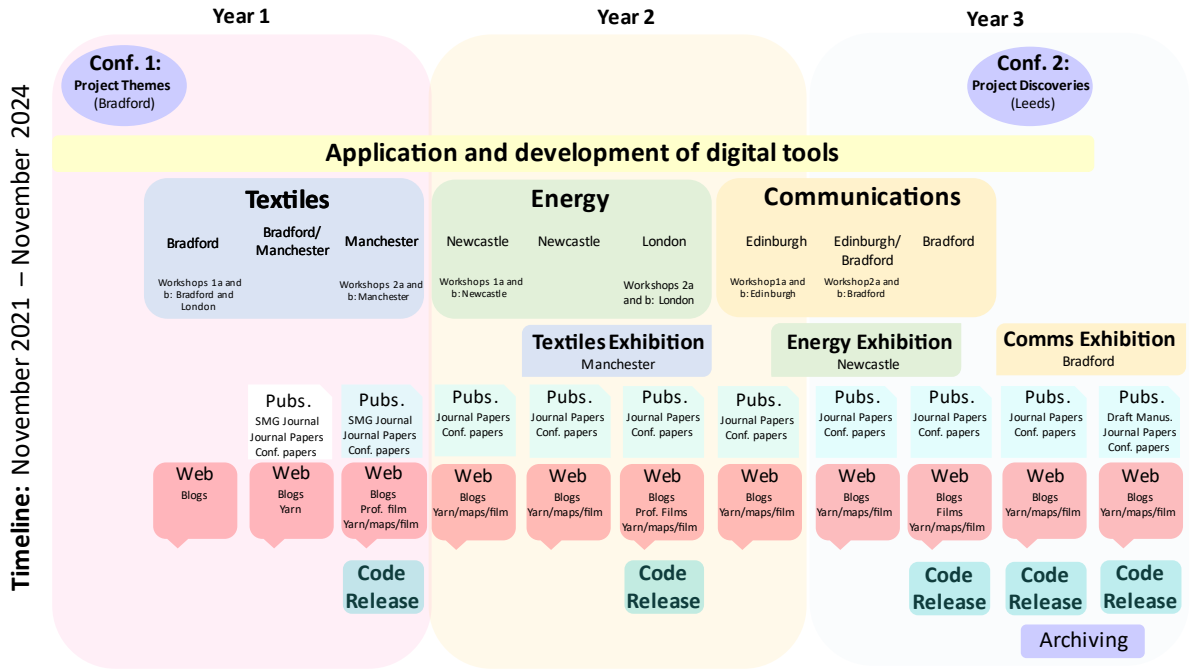
Historical / curatorial WPs 6-8:

- How does attention to assemblages of data on heterogeneous heritage items – objects, pictures, archives, films, TV and radio programmes, recordings, buildings, maps, places, etc – treated as historical sources – affect the historical narratives that historians, communities and curators may construct?
    - Textiles: Will Ashworth; Research Fellow, Stefania Zardini-Lacedelli; associated museum: Bradford Industrial Museum (Vicky Shaw, Lauren Padgett)
    - Energy: Graeme Gooday; Research Fellow now needing to be recruited; associated museum: Discovery Museum Newcastle (Kylea Little)
    - Communications: Jon Agar; Research Fellow, Daniel Wilson; associated museum: National Museums Scotland.

In addition to this formal structure, with our systemic action research methodology we have established a working group approach, under which team members form groups to pursue particular pieces of work; some of these are broad and will run across the project, such as the publishing working group ‘a forum for all aspects towards publishing, public speaking, communicating our work on the Congruence Engine project’. Others cover a specific deliverable and will be archived after delivery, such as the working group for planning the Science Museum Group Journal Special Issue, which will be published in December 2022. These working groups, which may equally be formed by research fellows, investigators or project partners, mobilise collaboration across the project, enabling the most effective contributions of enthusiasm, expertise and experience, whatever the individuals’ core role in the project. We are managing the project via the online ‘Basecamp’ Tool, where each of the working groups may be seen. As a result of our first ‘reflection’ workshop on 27th July, we simplified the structure, responding to feedback that it would be easier to use if it had fewer strands (screengrab, 24 August 2022):



Overall programme



Events and consultations

Date	Title / description	Type	Person's name	Approx. number reached	Audience type	URL
20-Oct-21	The Congruence Engine: Digital Tools for New Collections-Based Industrial Histories	Post-award Presentation	Tim Boon	50	SMG Collections Team	
28-Jan-22	The Congruence Engine: Digital Tools for New Collections-Based Industrial Histories: SMG All Staff Briefing	Event	Tim Boon	300	SMG All Staff	
Dec-2021	Project Webpage Created	Website		679 visits (February – July 2022)	General	https://www.sciencemuseumgroup.org.uk/project/the-congruence-engine/
Jan-2022	Project Blog Created	Website			General	https://ceblog.sciencemuseumgroup.org.uk/
9-11-Feb-22	Congruence Engine Opening Conference	Event		50	Co-Is, Project Partners, Project Board, Steering Committee	
Mar-22	Project Basecamp	Website		50	Co-Is, Project Partners, Project Board, Steering Committee	https://3.basecamp.com/5316423/projects
20-21-Jun-22	Textiles Planning Workshop (Leeds)	Workshop		15	Co-Is and Project Partners for Textiles, CE SMG Tean	The textiles pilot workshop, 20/21 June, Leeds (basecamp.com)
22-Jul-22	The Congruence Engine: Digital Tools for New Collections-Based Industrial Histories: Session at Annual Conference of the British Society for the History of Science, Belfast	Event	Tim Boon	40	BSHS delegates: historians of science and technology	

27-Jul-22	Textiles Reflection and Planning Workshop (London)	Workshop		15	Co-Is and Project Partners for Textiles, CE SMG Tean	
19-Aug-22	Panel discussion contribution on CE to 'Industrial Labour and Cultural Engagement in the Long 19th Century' conference of the 'Piston, Press and Pen AHRC project.	Conference	Tim Boon	30	Historians of 19thc literature and history	https://www.eventbrite.co.uk/e/industrial-labour-and-cultural-engagement-in-the-long-19th-century-online-tickets-377679578787

Research Approach

The Congruence Engine is being run within a systemic action research approach, one that is designed to enable investigation of complexity. Like many participatory projects delivered under the AHRC-led 'Connected Communities' theme, *Congruence Engine* includes multiple participants. It is the systemic action research methodology that makes this possible. CI Graham has successfully run two substantial AHRC projects, *Heritage Decisions* (with Boon as CI, AH/K006754/1) and *Bradford's National Museum* (with SMG Science and Media Museum, AH/P008585/1) using this approach. Here we are influenced by Danny Burns' *Systemic Action Research* (2007), which offers practical ways to design participatory research at scale to enable the different parallel enquiries and to structure crosscutting events to build on the insights they produce. The central philosophy of this approach is to research 'with' rather than 'on' people in ways that value practical as well as theoretical knowledge – 'the creative action of people to address issues that matter to them' (Heron and Reason 2001). Its design principles are: Multiple perspectives – we cannot understand the complexity of barriers to accessing and using collections only from the academic perspective; we must allow for a plurality of motivations and experiences (Burns 2007); Value different ways of knowing – all are intrinsically valuable; in our endeavour we need the experiential and affective as much as specific empirical knowledge (Heron & Reason 2001); Learn together through doing – we cannot understand challenges to establishing a national collection only in a theoretical way; we need to try, test and push the systems that already exist to work out how they can change (Bradbury *et al* 2019).

The action research methodology, as well as enabling the research, collectively convenes the participants to collaboratively trace, document and develop the project, evaluating barriers encountered and, at the highest level, interrogates the value of the project aim.

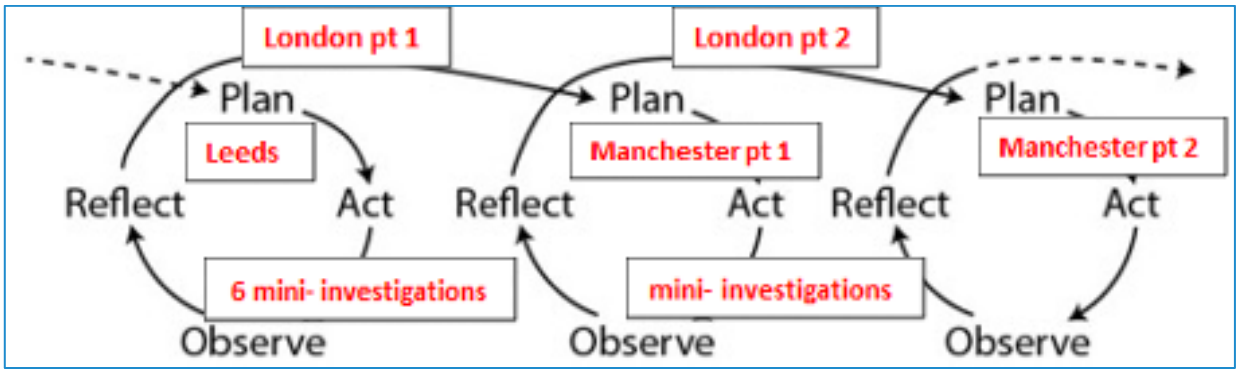
Progress

We have passed through the project setup phase, which began from project start in mid-November 2021 until February 2022. That phase was devoted to staff appointments, establishing initial meetings sequences (management / management and Action Research; Whole-Team; Project Board (integration with SMG systems); establishment of budgetary systems in alignment with Science Museum Group conventions. We have subsequently adjusted the meetings regime to ensure it is doing the right work for the project, notably by integrating the project management meeting with the main strand of research at any one time (currently textiles).

Since the 'We Are Congruence Engine' Opening Conference held in Saltaire and Bradford, 9–11 February 2022, we have been conducting the research along two critical paths:

1. We have been running a 'textiles pilot'. In this, we have embarked on the historical investigation of the first of our three industrial sectors, textiles, in the TaNC spirit of bringing together data from heterogeneous collections so that historians and curators – both professional and amateur – can explore the historical possibilities of cross-collection work. At the same time, we have been piloting our action-research model of participation and its co-production of the investigation. We held a planning workshop on 20-21 Leeds in June that brought together co-investigators, researchers and participants in the co-production of a series of mini-investigations that were then conducted in the weeks up to the reflection workshop in London on 27 July. In the action research cycle, the first

meeting represented the first 'plan' stage; the six mini- investigations combined 'act' and 'observe;' the later workshop achieved the 'reflection' and second 'planning' phase, as we move into the concluding part of the textiles investigation.



Annotated illustration of a basic action research cycle (Ejbye-Ernst, D. and Jørring, N.T., 2017: 53).

2. In parallel, we have one other main critical path for activity in 2022: delivery of a special issue of the Science Museum Group Journal devoted entirely to the project. This is not simply reportage of project activity; rather it is a major articulation of the project members' engagement with the TaNC and *Congruence Engine* problematics. See Project Outputs below for detail.

Although the core research is at a very early stage, we are able to begin to discuss how we are addressing the TaNC Impact Areas:

A. Explore how thousands of disparate collections could be explored by public audiences and academic researchers in the future.

The pilot study we have been conducting since February has been testing our assumptions on how participation should work within the project, and this has been done on the territory of our investigation into textile collections and history by means of conducting mini-investigations. At this stage, the 'audience' of four participating individuals from the Saltaire History Club and Collection, along with other project associates, is very small indeed, but the individuals we are working with are engaged as equals (alongside university academics, museum professionals and project researchers) in the investigation. In the second half of the textiles investigation, we are circulating a call to a wider range of potential participant groups, invited to shape, and/or to participate in the next round of mini-investigations.

This phase has deliberately started in a focussed way, looking into how the histories of Salts Mill (Saltaire) and Listers Mill (Manningham) can be reinfused with their social history using collections-linking software techniques. We have brought together several hundred records from a diversity of project partners to begin to model what linking collections might mean to the Saltaire History Club, and to the humanities and curatorial and the digital researchers within the project.

B. Dissolve barriers between collections and open up public access.

These must be taken separately. The pilot work is investigating how we may dissolve barriers between collections; it is modelling at the very small scale the affordances of opening up public access by co-

producing some techniques to do so. At the same time, it is probing the demand for such cross-collections enquiry. At this project stage we cannot open up public access more generally; exploring what it would mean to do so is an emerging research question, anticipated in the grant application's sub-research question under the umbrella heading 'what new opportunities might arise for the practice of history and for the museum-visiting general public from a national collection realised in the ways explored in this project?'

When we stage our first exhibit (target: spring 2023), we will begin to communicate some of this potential to museum visiting audiences, as we are already for the online audience with our project blog, and as we will also do more extensively in the planned special issue of our Journal this winter. The effectiveness and reach of the exhibits will be evaluated.

C. Set a global standard for other countries building their own collections. And D: Enhance collaboration between UK and national collections worldwide.

We plan to enrol international audiences via conferences, including the Artefacts Consortium (October) and the History of Science Society (November), and via our online *Science Museum Group Journal*.

In terms of Key UX/Engagement goals under the FAIR principles of Findability, Accessibility, Interoperability, Reusability; these are being explored in the texture of the investigation, and there will be more to say in the midpoint and final reports. Our public engagement classification of audiences includes professional and lay researchers and curators. Our social-technical approach of systemic action research socially organises our execution of historical and digital work. The approach is intrinsically bottom-up, responsive to participant audience needs.

Research Results

Some tentative findings are already emerging, and we will be investigating these as the project develops. The action research approach has enabled some shifts in response to both macro- and micro- factors. At the small scale, we have observed how the 'material conditions' of the project (especially the differential amounts of time that different participants are funded to be on the project and its distributed nature) places a premium on the full-time staff (Research Fellows) taking a larger hand in the mini-investigations than we had anticipated. At the scale of the whole set of TaNC Discovery projects, *Congruence Engine* is developing a strong identity around techniques for the collaborative co-production of the national industrial history collection; the investigation is beginning to enact the work of creating a national collection by experimental acts of linkage. We have tentatively described this as thinking of 'to national collection' as a verb, denoting the sense that the national collection is not a noun – something that already exists awaiting linkage – but something that is created by means of co-production using digital techniques. Conceived in this way, the 'national collection' becomes like a fractal: at every level of increased magnification, there is more detail below.

Our work entails a shift of weight away from what might have been a core technique – the use of machine learning techniques to create an aggregated knowledge graph of industrial collections, which is similar to the activity of three of the other Discovery Projects (and indeed to *Heritage Connector*) – towards a greater emphasis on the pragmatic use of off-the-peg digital humanities tools. This is partially because, at least at the start of the project, the participants' sense of what it might mean to link collections is predetermined by existing historical practice, which tends to be at the 'hand-stitched' level of drawing on a very few historical sources; we are regularly interrogating the scope for scale-up to bigger data and approaches using machine learning, where such approaches can serve the research needs of making the national industrial history collection. Associated with all of this has been a move towards a more balanced co-production model, where we have substituted an approach in which we had assumed that the enquiries of historians and curators would be serviced by digital experts with one in which participants from any background, digital as well as curatorial or historical, can suggest mini-investigations. The emergence of this factor is linked to the next point:

a) Cross-collections research is, for many, a novel proposition

There are two related unstated assumptions in the TaNC enterprise. The first is that, if they could, people would want to be able to work across collections. At the simpler level, this might be envisaged as a search engine for UK collections, so that it would be possible to find, for example, all the artefacts of one kind in British collections, or everything associated with a single individual. Such a facility would undoubtedly be a boon to researchers, for all that its achievement with existing thin and inconsistent collections data presents a major technical challenge. We provided proof of concept in our Foundation Project, *Heritage Connector*, that this is possible by constructing knowledge graphs, using machine learning techniques to link together key terms within the graph, including items from Wikidata. Issues of scalability to the national level and sustainability of the results have yet to be investigated, and they are not a major focus of *Congruence Engine.*

b) Linking collections for new kinds of historical research

Congruence Engine is much more concerned with the other untested assumption, namely the potential of linking collections for new kinds of historical and curatorial practice, both by people who are paid to do

history, and those who do it for fun. The evidence from our first research phase is that this is a novel proposition for many of the individuals and groups we are working with. Family and local history researchers know how to work with the census, for example, these days online via services such as Ancestry. Virtually all historians become skilled in working with paper archives, and then perhaps turn to picture archives for illustrations for their presentations and publications. But using a combination of different kinds of collections as forms of evidence of history is a new idea for many – with the possible exception of museum curators, who are used to bringing together media when they create exhibitions. But even they do not systematically publish in conventional textual forms. So, *Congruence Engine* is exercising a measure of vanguardism in seeking to explore what research will be like when it is possible to work across collections in a more systematic way than is currently easily possible.

c) Affordances of different kinds of heritage in creating a national collection

At this point, the priorities of the participants have focussed on kinds of social-historical possibility. They want to be able to discover more about the lives of ordinary people who lived in Saltaire or Manningham. But the kinds of data associated with, for example, museum collections of textile machinery do not lend themselves readily to enabling the 'history from below' of ordinary lives lived in those places; where the data include associated names, they tend to be of inventors and makers, not machine hands. But this is where research undertaken by associates of the Saltaire History Collection using bigger data sources and other available listings have been very suggestive. For example, Colin Coates' analysis of Saltaire Census data 1851-1921 makes it possible to explore patterns of migration and kinship in Saltaire, whilst a classified list of textile trades derived from the 1921 census makes it possible to link named individuals to particular trades, and therefore machines, and thereby particular locations within Salts Mill. In other words, these data-heavy documents, themselves parts of the digital national collection, act as a kind of 'connective tissue' enabling the linking of records of material things such as objects and pictures, and potentially of collections of all kinds.

d) Appreciating multi/interdisciplinary cultural differences

At the outset, we realised that there exist significant differences in the languages used by the four main working cultures – historical and curatorial (both professional and amateur), digital, and action research - that we need to work together for successful delivery of the project. It has become clear that these linguistic differences also reflect differing working cultures, practices and expectations. We believe that these are worth investigating, surfacing and discussing because with greater understanding we should enable more effective collaborative working. With this in view, we have begun discussions with social scientists whom we are inviting to work with us.

Project Outputs

The Project Blog has so far featured:

- We are Congruence Engine: Metaphors and Project Conduct (22 Feb)
- Methods: Seeking congruence and enabling divergence (14 March)
- The Congruence Engine opening conference – enacting the principles of systemic action research (24 March)
- Congruence Engine North Star (24 March)
- Energising materials, connecting stories – local, national and international (25 May)
- What can Omeka do for your digital journey? Reflections from the first Congruence Engine Pilot Study (1 Jul)
- Co-producing research inquiries for the textiles strand (2 Aug)
- Reflecting on the Textiles Pilot (22 Aug)

Magazine Publications

Rees, Arran. opinion piece, *Museums Journal* (Museum Association Magazine), Jan/Feb 2022, pp 12., https://www.museumsassociation.org/museums-journal/opinion/2022/01/digital-the-potential-of-ai/

Boon, Tim. "The Congruence Engine." *Viewpoint* (BSHS Magazine), no. 126, Feb. 2022, pp. 8–9. https://www.bshs.org.uk/wp-content/uploads/No-126-Feb-2022-for-web.pdf

Congruence Engine special issue of the Science Museum Group Journal

We plan to publish this issue in December 2022. This involves a remarkably tight turnaround for a refereed journal, but this is possible because of the SMGJ's experienced, nimble production methodology and the retained staff at the Science Museum. We have a good list of propositions, from the project team mainly:

Author	Article	Article type/subject/origin
Introduction/Editorial		
Tim Boon	Origins and Ambitions of the Congruence Engine Project	Single author article
Helen Graham, Arran Rees	Writing, sharing, feeling, doing: action research and the Congruence Engine	Editorial
Research papers		
Will Ashworth	First thoughts from other textile/industrial history researchers?	Single author article
Paul Craddock	Essay film on Saltaire experiment	Film essay
Stefania Zardini	Crossing the boundaries: the experimental value of digital platforms in connecting and reimagining heritage collections	Single author article

Tim Smith	Congruence Engine: Woollen Textiles Photo Essay	Photo essay
John Stack, Jamie Unwin	The potential and pitfalls of machine learning in the Congruence Engine context (or other)	Two-author article
Graeme Gooday, Kylea Little, Cameron Tailford	Energizing Connections in Museum Collections	Multi-author article
Asa Calow	(on machine learning)	
Alex Butterworth	(on digital-historical ontologies)	Single author opinion piece
Ross Parry	Congruence Engine, TaNC and the state of the art in GLAM digital practice	Single author article
Discussion papers/Conversations		
Stefania Zardini Lacedelli, Anna-Maria Sichani, Arran Rees	Keywords for the Congruence Engine (subject to review after death of Research Fellow Cameron Tailford)	Conversation
Stuart Prior	How to get from local knowledge to global knowledge: the barriers between people and Wikipedia	Single author article
Kylea Little, Ellie Swinbank, Felicity McWilliams	The Making of Industrial Collections in Edinburgh, Newcastle and Birmingham	Discussion paper
Wayne Cocroft	Connecting places and collections	Hybrid: discussion / photo essay
Patrick Russell and Tim Boon	Congruence Engine: Implications and Possibilities for the Study of Industrial Films	Co-Written
Jon Agar	Communications and the Congruence Engine: Early Thoughts and Possibilities	Single author article
Daniel Wilson	Working at Scale; or, what do digital methods mean for research using cases, models and collections?	Single author article
Simon Popple and Stefania Zardini Lacedelli in conversation with Stuart Prior, Maggie Smith, Arran Rees.	Building Community stories and exploring emerging narratives inside Congruence Engine	Conversation piece
Jane Winters and Anna-Maria Sichani	The role of the Digital Humanities in an interdisciplinary research project	Two-author article

Cross-project Collaboration

Three of the other Discovery Projects have agreed to referee submissions to the special issue of the *Science Museum Group Journal.* At the 14th September 2022 meeting I also propose to see whether other Discovery Projects might be interested in a cross-project study of interdisciplinary collaborations, and how to resolve any issues arising from the working cultures from them.

Sustainability and Infrastructure

The project will be working with various cross-content types of input data:

1. data owned by the Science Museum Group;
2. other partner cultural heritage institutions supplied;
3. Wikidata

- A project-led GitHub (public) repository, under the project's GitHub account is currently being employed to ensure secure storage, open, inclusive access and short-term maintenance of the cross-content datasets we will be ingesting from project partners throughout the project. These datasets will be stored as raw/master data files, ready to be (re)used across the various project's mini-investigations, with built-in version control and advanced collaborative functionalities (wiki). A private GitHub Repository will be used for datasets that are lacking open licences.
- Under designated repositories for the project's strands (textiles, energy, communications), we are also storing all the newly developed code for the project's prototypes and investigations, alongside the project's public-facing documentation and reports (in line with guidelines being developed by the Publishing Working Group) and training material, as produced and used throughout the project.
- Project prototypes, data-related outputs and project-specific updates will be hosted in a dedicated-SMG website https://www.sciencemuseumgroup.org.uk/project/the-congruence engine/ ensuring continued access and (re)use of the project's digital outputs. This will be the central place for the project to showcase and communicate its findings in an accessible way, following also TaNC-AHRC requirements for publicly accessible outputs.
- A Zotero library has been set up, with a number of sub libraries, to store the project bibliography.
- Basecamp, currently hosted by MadLab (500GB storage included in their fee), is used for day-to-day communication among team members, general storage, updates and networking. A backup strategy for its contents is being developed.

Licensing and Copyright

Project partners' datasets are normally shared with the project under open licences such as CC0 - public domain data or CC-BY-SA (Attribution-Share Alike). Under these licences, data can be further processed and openly published in our *Congruence Engine* Github repository. Where the data provider specifically claims a different copyright for their data, their dataset will remain in the **private Github repository.** Licensing will be made clear for all datasets and be explicitly stated in the dataset documentation (see section below).

Data documentation

A developing set of documentation guidelines is also in place for all datasets we are currently and will be using, containing the structured information shown below, in line with the Datasheets for Datasets (Gebru et al 2021) proposal:

- Description
- Copyright
- Source URL
- Use cases

- Linked datasets
- Potential datasets to use alongside this dataset

Known History / things to be aware of

- Example projects using this data
- Known Biases
- Known History / Things to be aware of
- Is this a derivative or intermediate dataset?
- Can it be regenerated from other datasets or sources elsewhere? If so how? Life of dataset / Can or should it ever be deleted
- Owner(s)

Where appropriate, we will provide Persistent Identifiers (PID) for all stored and (re)used datasets as well as for project outputs.

Contacts

Tim Boon	PI	tim.boon@sciencemuseum.ac.uk	Overall
Nina Webb-Bourne	Coordinator	Katerina.Webb-Bourne@ScienceMuseum.ac.uk	Practicalities
John Stack	Digital Co-I	John.Stack@ScienceMuseum.ac.uk	Digital
Jane Winters	DH Co-I	jane.winters@sas.ac.uk	Digital Humanities
Anna-Maria Sichani	DH Research Fellow	annamaria.sichani@sas.ac.uk	Digital Humanities
Helen Graham	AR Co-I	h.graham@leeds.ac.uk	Action Research
Arran Rees	AR Research Fellow	a.j.rees@leeds.ac.uk	Action Research
Will Ashworth	Textiles Co-I	w.j.ashworth@liverpool.ac.uk	Textiles History
Graeme Gooday	Energy Co-I	g.j.n.gooday@leeds.ac.uk	Energy History
Jon Agar	Communications Co-I	jonathan.agar@ucl.ac.uk	Communications History

Bibliography

Bradbury, Hilary, Steve Waddell, Karen O' Brien, Marina Apgar, Ben Teehankee, and Ioan Fazey, 'A Call to Action Research for Transformations: The Times Demand It', *Action Research*, 17.1 (2019), 3–10 <https://doi.org/10.1177/1476750319829633>

Burns, Danny, *Systemic Action Research: A Strategy for Whole System Change* (Bristol: Policy Press, 2007)

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and others, 'Datasheets for Datasets', *Communications of the ACM*, 64.12 (2021), 86–92 <https://doi.org/10.1145/3458723>

Heron, John, and Peter Reason, 'The Practice of Co-Operative Inquiry: Research with Rather than on People', in *Handbook of Action Research: Participative Inquiry and Practice*, ed. by Peter Reason and Hilary Bradbury (SAGE Publications, 2001)