Museum online catalogues-as-data | report

A Congruence Engine mini-investigation (January – June 2023)

Dr Anna-Maria Sichani, Jamie Unwin
(Critical reader: John Stack
Assistance with data/code: Kunika Kono)

Executive summary

This report, output of the investigation titled **Museum online catalogues-as-data,** aims to contribute towards the design and development of a **sector-level strategy regarding the improvement and connectivity of online catalogues** of cultural heritage institutions as part of a national collection's infrastructure. Through the investigation we embark on exploring and demystifying common practices in museum online catalogues, aiming to reveal potential obstacles that suspend their connectivity and advanced use from various users.
Cultural heritage institutions' and more specifically museums' online catalogues are rich information architectures, usually including digitised versions/surrogates of the analogue assets alongside associated catalogue metadata in various formats, combined with data management and discovery tools. Although increasingly large focus and investment in the digital cultural heritage sector is given into advanced computing power and functionalities of online catalogues as well as interface design, **the 'connecting tissue'** that will allow connectivity between online museum catalogues, and thus collections, is still a *desideratum.*
The first part of the report offers **a state-of-the-art of digital cataloguing practices** in the museum sector. The second part focuses on specific technical challenges, found in online museum catalogues, by introducing a **sector-specific assessment framework** through a combination of empirical research and landscape analysis. For this framework we employed Science Museum Group (SMG) online collection catalogue as a case study to showcase existing challenges and limitations. Finally, in the third part **a minimum technical passive provision** is explored in order to serve as a foundational infrastructure for drawing together a cultural heritage national collection. The Appendix hosts two questions-as-provocations that we invite the Congruence Engine team to consider further.

Introduction

In the field of digital cultural heritage, museum collections are often seen as 'ideal' datasets for computationally driven research and work. Not surprisingly, museum online catalogues are thus used as a public-facing entry point to museum collections, enhancing collections' discovery, access and exploration while ensuring engagement with various audiences.

Cultural heritage institutions' and more specifically museums' online catalogues are rich information architectures, usually including digitised versions/surrogates of the analogue assets alongside associated catalogue metadata in various formats, combined with data management and discovery tools. Although increasingly large focus and investment in the digital cultural heritage sector is given into advanced computing power and functionalities of online catalogues as well as interface design, **the 'connecting tissue'** that will allow connectivity between online museum catalogues, and thus collections, is still a *desideratum.* Towards a National Collection's objectives are also highlighting the importance of connecting while opening up collections from the outset:

- to begin to dissolve barriers between different collections
- to open up collections to new cross-disciplinary and cross-collection lines of research
- to extend researcher and public access beyond the physical boundaries of their location
- to benefit a diverse range of audiences
- to be active and of benefit across the UK
- to provide clear evidence and exemplars that support enhanced funding going forward.

'Museum online catalogues-as-data' takes inspiration from the recent GLAM initiative 'Collections-as-data' (Padilla et al. 2019) and focuses on ways for enabling better (re)use and connection between museum online catalogues and collections. In order to better approach our goal, we seek to answer the following questions:

- How easily and efficiently can we access, use and link museums' collections via their online catalogues?
- What are the obstacles and weaknesses of the catalogues' underlying data that limits this linking and how can we address these?
- In which ways do current and emerging systems, technologies and practices support the creative reuse and connectivity of museum online catalogues and their records?

Goals

This investigation aims to contribute towards the design and development of a **sector-level strategy regarding the optimization, improvement and connectivity of online catalogues** of cultural heritage institutions as part of a national collection's infrastructure. Through this investigation we embark on exploring and demystifying common practices in museum online catalogues, aiming to reveal potential obstacles that suspend their connectivity and advanced use from various users.

Methodology

The first part offers **a state-of-the-art of digital cataloguing practices** in the museum sector and the second part focuses on specific technical challenges, found in online museum catalogues, by introducing a **sector-specific assessment framework** through a combination of empirical research and landscape analysis, by using existing data collected mainly via TaNC foundation projects and consolidated reports as well as other resources produced by stakeholders. For this framework we employed Science Museum Group (SMG) online collection catalogue as a case study to showcase existing solutions and limitations[1].

[1] In 2017 the Science Museum Group relaunched its online collection website. The website consolidated a number of existing websites publishing digitised collection material — organised by subject or area of collection — into a single presence.

Finally, by adopting a 'good-enough' ethos in our practice[2], **a minimum technical passive provision** is explored in order to serve as a foundational infrastructure for drawing together the national collection.

> Such an attempt, although emergent and incomplete, should be seen as a first step towards an empirical assessment and realistic thinking exercise of what are we are missing and/or how might we think and work differently while designing the infrastructure for a National Collection.
> Grounding the "connecting is good " with a bit of "connecting is good because it enables X and Y" is probably worth considering as it frames the whole endeavour of the Congruence Engine (CE) as part of the Towards a National Collection program, especially within the framework of a social machine. Especially, as CE investigations are by design highly experimental ones but might feel disjointed at times, it would be good if together they formed examples within a "connecting/linking collections enables X and X is good for Y" framework, even if that framework is still a work-in-progress.

A. Digital cataloguing practices: a state-of-the-art

Over the last twenty years of museums' digitisation boom, institutions have moved forward with new technologies and workflows for the (internal) documentation and cataloguing of their records, as well as for their online publication and discovery (Luther 2022).
Most cultural heritage institutions, as part of their institutional practice, have been active in documenting the standards (eg. ISAD(G), Spectrum), the internal digital cataloguing and management system(s), their processes and technical choices throughout the years, often communicating openly their decisions and overall strategy (see e.g. for the UK The TNA Digital cataloguing practices report (2017). In addition, there has been a recent set of initiatives on historicising and assessing museum (digital) cataloguing practices, including formative and summative evaluations of online catalogues. The Digital Catalogue study. A cross-institutional user study of online museum collection catalogues (November 2019) focuses on digital catalogues published by the Art Institute of Chicago, the J. Paul Getty Museum, the National Gallery of Art (USA), and the Philadelphia Museum of Art, by focusing on a number of user-experience parameters such as Marketing and Demographics, Functionality and Design, Scholarly Content and Measuring Success. Under the AHRC-funded research programme Towards a National Collection (TaNC), the Digital collections Audit was carried out by the Collections Trust, between September 2021 and the end of January 2022, focusing on the specific attributes of digitally accessible collections of UK's leading cultural heritage institutions, including their digital catalogues, aiming to support the development of a future national digital collection infrastructure.
In addition, there have been also usability studies and reports on user needs with a focus on digital collections and catalogues for cultural heritage institutions, like the Online User Research Literature Review by Dr Claire Bailey-Ross (2021) as well as the User Research by The Audience Agency and Culture 24 (2022), both commissioned by TaNC.
Finally, a recent comprehensive report by Collections Trust entitled Getting it together: realising the value of museum collections data (2021) focuses on the value and

[2] See also Appendix B.

requirements of sharing UK museums collections data, with a dedicated section on online catalogues' requirements.

B. Assessing museum online catalogues

This section aims to introduce a twofold assessment framework for online museum catalogues.

Firstly, this framework is influenced by the "Five Stars of Open Data", a set of considerations for Publishing Data Online developed by Sir Tim Berners-Lee, inventor of the World Wide Web.

* data must be available on the Web under an open licence
** data must be in the form of structured data
*** data must be in a non-proprietary file format
**** data must use URIs as its identifiers , so can people point to it
***** data must include links to other data sources

For an application of the five stars system to museum online catalogues:

- **One star** describes most online catalogues on museum websites as the data often only allows a **read-only access,** with no provision for being automatically downloaded or retrieved.
- **Two stars** describes data publishing as structured data, including images and other digital material as separate files.
- **Three stars** describes all of the above and data that is made accessible in repositories with data that is parsed into manageable files that have clear non-proprietary format (e.g. CSV instead of Excel). This data might be stored in Relational Databases (RDB) and could also be accessible in APIs.
- **Four stars** describe data that is available using a Linked Open Data format (URI), under an open licence allowing them to be connected to central storage repositories such as Wikidata.
- **Five stars** can be described as connections that an institution makes to a larger context outside of the GLAM context.

Secondly, our assessment framework is based on basic functions online museum catalogues are designed to offer on/for collections' data:

1. Structure
2. Access and (Re)use
3. Link
4. Discovery and Search

Below you can find a matrix of these two approaches for assessing online museum catalogues:

Star Level	Description	Functionality

1	Online catalogue where data only allows read-only **access,** with no provision for being automatically downloaded or retrieved	(Basic) access (Basic) search
2	Data published as structured data including images and other digital material as separate files	(Basic) structure (Basic) access (Basic) search
3	Data made accessible in machine-readable form using a non-proprietary format. This data might be stored in Relational Databases (RDB) and could also be accessible in APIs.	Advanced structure Access (Re)use Search
4	Data made available using a Linked Open Data format (URI), under an open licence allowing them to be connected to central knowledge repositories (eg Wikidata).	Advanced structure Access (Re)use Link Search
5	Connections that an institution makes to a larger context outside of the GLAM context.	Advanced structure Access (Re)use Link Search Discovery

1. ‘Gremlins’ in the catalogue data: data structure and documentation in legacy cataloguing systems

Museum catalogues are built upon decades of museum collection and documentation practices, much of it in a predating online access and used to manage the collection rather than for public access. In other words, museum online catalogues are the last chapter of a rather complicated history of documentation and cataloguing practices and systems, set aside digitisation attempts, produced and employed by various subjects with different knowledge and expertise through the years, which often slips unnoticed when we are embarking on large-scale computational work within or among different online catalogues and collections[3].

A recent TaNC audit on digital collections revealed that many UK cultural heritage institutions are “dissatisfied with current systems and processes. In most cases, this was down to frustrations with legacy systems that didn’t have APIs resulting in difficulties integrating with other systems and services. A large number of institutions stated that they had either just embarked on the process of implementing a new collections management or digital asset management system or had budget approval to start the process. Several institutions acknowledged that they were in a state of flux, knowing that the transition from current to new systems would take several years” (Gosling et al 2022, 26).

Museum catalogues, thus, are more than flat and homogeneous data inventories; they often contain information compiled by many people over many decades, containing layers of historical/previous data documentation approaches thus resulting in huge diversity and disparity in data across the collection. Data documentation and information management, alongside curatorship, have been at the forefront of museum practice since the early days bringing together assets, processes, subjects and systems. From index cards to early database systems and then to recent data management systems, as cataloguing systems and metadata standards change and technology evolves, information about museum assets changes too, becomes layered, overwritten, enriched, duplicated, often gets lost or omitted. Curatorial decisions, data transition and migration, system upgrades and platform changes, staff changes, all lead to some sort of legacy information changes in catalogue records, without much background documentation about these changes.

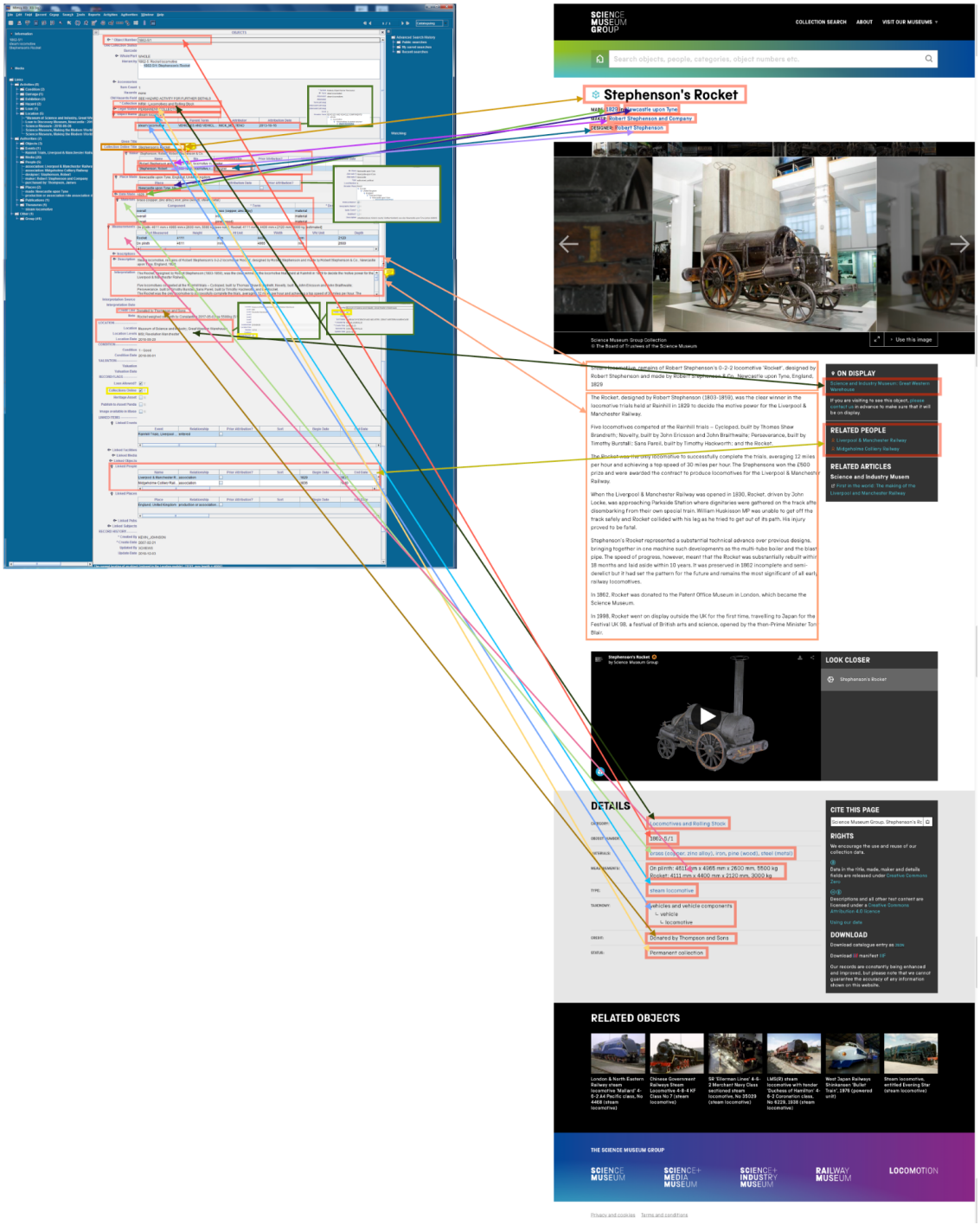
Not surprisingly, what we have in front of us today, in the form of online catalogues, is an accumulation of – or even an arbitrary choice from – the information compiled and stored on museum assets throughout the decades of collection documentation practice. Also, as the internal management system and the online collection database of a museum can entail different data depending on the level of editing, curation and staff that works, edits, cleans or vets information before they are published online, often the full records of museum collections via their catalogues are only accessible to internal staff. In addition, different museums might catalogue the same objects differently: there are fields which are authoritative in the catalogue (e.g. manufacturer) and others which exist in the context of

[3] see the AHRC-funded project “Legacies of Catalogue Descriptions and Curatorial Voice: Opportunities for Digital Scholarship” https://cataloguelegacies.github.io , https://lucyhavens.com/legacies-of-catalogues-infographics

collecting/collections for the owning organisation (e.g. description), a practice that impedes automated linking. Museums may make some (although rarely all) of their records publicly available, but this is usually just the catalogue entry, with the ‘public access’ fields in the catalogue systems in mind (such as ‘web description), simply republished online.

These legacies of cataloguing practices and records’ documentation explain why museums’ datasets are far from a neat, perfect set of structured data, thus impeding advanced analytical procedures and linking between collections. We should also acknowledge that it’s part of cultural heritage institutions’ thinking of the catalogue as only worthy of publication when it is “complete” or "ready" and so there can be a reluctance to publish online records perceived as being incomplete. One could argue that since these catalogues are forever a work in progress, a bolder approach could be taken.

An interesting example from the Science Museum Group is the mapping exercise between the Mimsy entry fields (the collection management system used for objects, whereas Adlib is used for archival collections and Koha for library collections) and the online catalogue fields, showcasing that not all catalogue data ends up in the online publicly available catalogue.

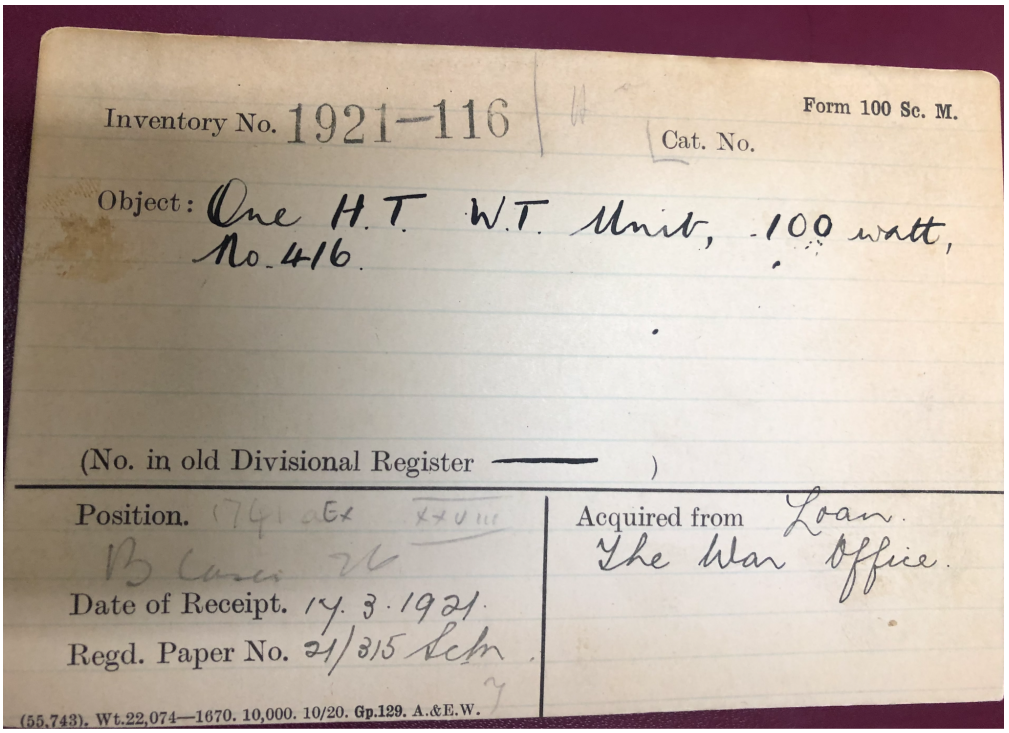


An SMG object biography through cataloguing instances

If we go one step back in the SMG cataloguing history[4], before the Mimsy entry, to the archival hard-copy record of the object (a) in question ('index cards'), we can notice that details of how the object came into the collection is often mixed in with personal data and commentaries of different agents through the years (for a more detailed and wider view on

[4] A separate report on SMG legacy cataloguing practices will also be produced to further support this assessment.

an object's trajectory within the museum collections, one can consult the SMG 'Green Files') While this data tells a fuller story, those records are not published online (b), both because it is hard to separate out what is private/personal and what is public and more mainly due to the cost of digitisation of SMG archival records. The data available through the API (c) is a combination of the Mimsy entry and the online version of the record.



Inventory No. 1921–116 Cat. No. Form 100 Sc. M.

Object: One H.T. W.T. Unit, 100 watt, No. 416.

(No. in old Divisional Register ———)

Position. 174 aEx XXVIII B Case 24

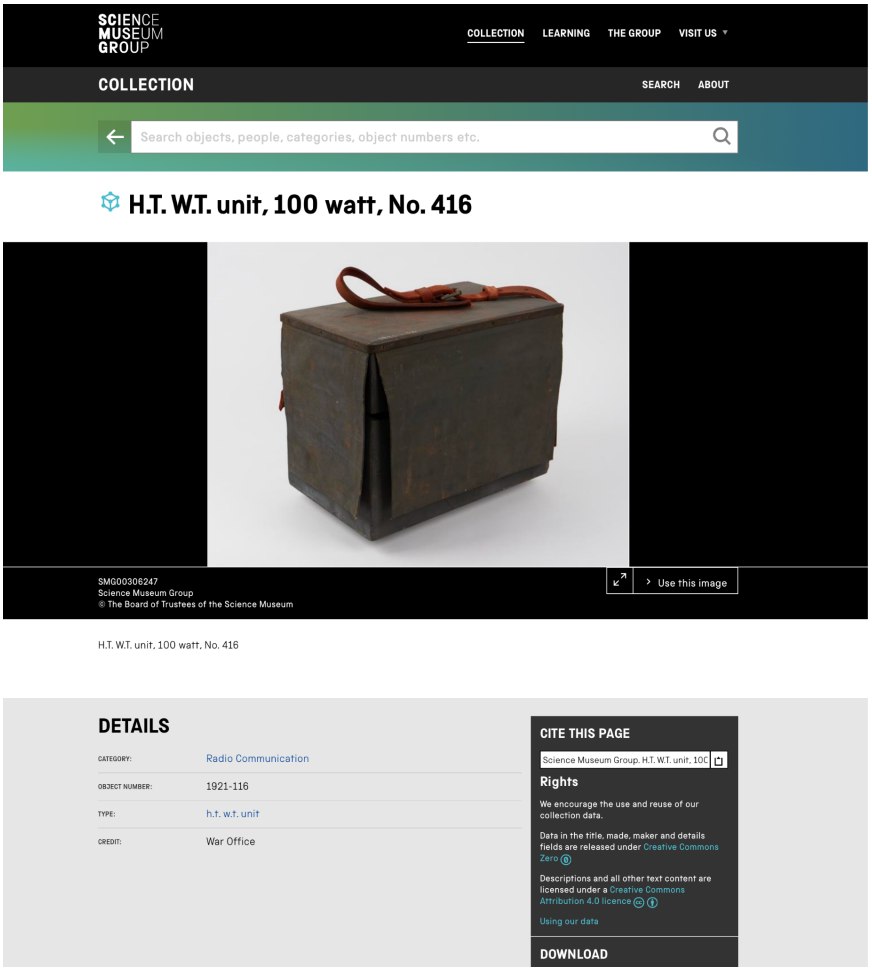
Acquired from Loan. The War Office.

Date of Receipt. 17. 3. 1921.

Regd. Paper No. 21/315 Sch.

(55,743). Wt.22,074—1670. 10,000. 10/20. Gp.129. A.&E.W.

(a) A Science Museum 'Form 100' record for the object



(b) The online version of the object's record https://collection.sciencemuseumgroup.org.uk/objects/co34900/h-t-w-t-unit-100-watt-no-416-h-t-w-t-unit


```
{
  "data": {
    "___comment": "### WARNING ### - <enhancement> tags are experimental, please do not
aggregate",
    "type": "objects",
    "id": "co34900",
    "attributes": {
      "admin": {
        "added": 1495031307000,
        "created": 826588800000,
        "id": "object-34900",
        "language": "eng",
        "modified": 1681925905000,
        "previous_status": "valid and published",
        "processed": 1682696024806,
        "source": "smgc",
        "stream": "smg-online",
        "uid": "co34900",
        "uuid": "40f25ee1-8570-319a-9bc5-c08898f70f42",
        "version": 26
      },
      "description": [ { "primary": true, "value": "H.T. W.T. unit, 100 watt, No. 416" } ],
      "identifier": [ { "primary": true, "type": "accession number", "value": "1921-116" }
],
      "name": [ { "primary": true, "value": "h.t. w.t. unit" } ],
      "summary_title": "H.T. W.T. unit, 100 watt, No. 416 (H.T. W.T. unit)",
      "title": [
        {
          "primary": true,
          "type": "display title",
          "value": "H.T. W.T. unit, 100 watt, No. 416"
        }
      ],
```

```
      "type": { "base": "object", "record_type": "1" },
      "categories": [
        {
          "museum": "SCM",
          "name": "Radio Communication",
          "value": "SCM - Radio Communication"
        }
      ],
      "language": [ "eng" ],
      "legal": { "credit_line": "War Office" },
      "measurements": { "item_count": "1" },
      "multimedia": [
        {
          "@link": { "type": "reference" },
          "admin": {
            "id": "media-490941",
            "source": "smgi",
            "uid": "i490941",
            "uuid": "1ab589a6-197d-3569-b987-4a0439dd69a8"
          },
          "credit": "Science Museum Group",
          "for_sale": "0",
          "identifier": [ { "type": "iBase id", "value": "490941" } ],
          "priority": "0.5",
          "processed": {
            "large": {
              "format": "jpeg",
              "location":
"https://coimages.sciencemuseumgroup.org.uk/images/490/941/large_SMG00306247.jpg",
              "measurements": {
                "dimensions": [
                  { "dimension": "height", "units": "pixels", "value": 1186 },
                  { "dimension": "width", "units": "pixels", "value": 1536 }
                ]
              },
              "modified": 1612291281000,
              "resizable": true,
              "type": "image"
            },
            "large_thumbnail": {
              "format": "jpeg",
              "location":
"https://coimages.sciencemuseumgroup.org.uk/images/490/941/large_thumbnail_SMG00306247.jpg"
,
              "measurements": {
                "dimensions": [
                  { "dimension": "height", "units": "pixels", "value": 247 },
                  { "dimension": "width", "units": "pixels", "value": 320 }
                ]
              },
              "modified": 1612291281000,
              "resizable": true,
              "type": "image"
            },
            "medium": {
              "format": "jpeg",
              "location":
"https://coimages.sciencemuseumgroup.org.uk/images/490/941/medium_SMG00306247.jpg",
              "measurements": {
                "dimensions": [
                  { "dimension": "height", "units": "pixels", "value": 576 },
                  { "dimension": "width", "units": "pixels", "value": 746 }
                ]
              },
              "modified": 1612291281000,
              "resizable": true,
              "type": "image"
            },
            "medium_thumbnail": {
              "format": "jpeg",
              "location":
"https://coimages.sciencemuseumgroup.org.uk/images/490/941/medium_thumbnail_SMG00306247.jpg
",
              "measurements": {
                "dimensions": [
                  { "dimension": "height", "units": "pixels", "value": 148 },
                  { "dimension": "width", "units": "pixels", "value": 192 }
```

```
              ]
            },
            "modified": 1612291281000,
            "resizable": true,
            "type": "image"
          },
          "small_thumbnail": {
            "format": "jpeg",
            "location":
"https://coimages.sciencemuseumgroup.org.uk/images/490/941/small_thumbnail_SMG00306247.jpg"
,
            "measurements": {
              "dimensions": [
                { "dimension": "height", "units": "pixels", "value": 48 },
                { "dimension": "width", "units": "pixels", "value": 62 }
              ]
            },
            "modified": 1612291281000,
            "resizable": true,
            "type": "image"
          },
          "zoom": {
            "format": "IIIF",
            "location":
"https://zoom.sciencemuseumgroup.org.uk/iiif/2/490%2F941%2FSMG00306247.ptif",
            "measurements": {
              "dimensions": [
                { "dimension": "height", "units": "pixels", "value": 4657 },
                { "dimension": "width", "units": "pixels", "value": 6032 }
              ]
            },
            "modified": 1612291281000,
            "resizable": true,
            "type": "image"
          }
        },
        "public_view": "1",
        "publish": "1",
        "sort": "2021-02-02 18:40:40.0",
        "source": {
          "legal": {
            "rights": [
              { "details": "© The Board of Trustees of the Science Museum" }
            ]
          },
          "title": [
            { "type": "caption", "value": "SMG00306247" },
            { "type": "main title", "value": "SMG00306247" }
          ]
        },
        "type": { "base": "media", "type": "image" }
      }
    ],
    "numbers": { "NUMBER1": "31" }
  },
  "links": {
    "self":
"https://collection.sciencemuseumgroup.org.uk/objects/co34900/h-t-w-t-unit-100-watt-no-416-
h-t-w-t-unit",
    "root": "https://collection.sciencemuseumgroup.org.uk",
    "api": "https://collection.sciencemuseumgroup.org.uk/api/objects/co34900",
    "iiif": "https://collection.sciencemuseumgroup.org.uk/iiif/objects/co34900"
  }
 },
 "inProduction": true
}
```


(c) Here is the downloadable (API) version of this record

2. Access & (Re)use: Open Data in museums (APIs, IIIF, etc.)

While museum online catalogues offer a straightforward, accessible way to navigate through museums' collections data, it is through the OpenGLAM movement that

museums started to explore ways towards “opening” all of collections data for anyone to freely use, reuse, or distribute it. This is a brave and radical move from the browse-and-read-only aspect of catalogues (one and two stars, based on our framework) towards a creative interaction with heritage data. In this context, data refers not only to a digital image of a collection record but all of an object’s metadata or supporting information: what one can usually find via an online catalogue. The magic of open collections data, though, is that through various tools and approaches, all those individual bits of information that someone can see via an online catalogue entry can be packaged together and unpacked, visualised, disseminated and reused in new creative ways.

Earlier attempts to offer openly available museum datasets include the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) or as machine-readable data via a GitHub repository, such as the New York Public Library, the Carnegie Museum of Art, The Cooper Hewitt Smithsonian Design Museum, the Tate Collection, and the Minneapolis Institute of Arts. Over the last ten years, the most common way for a museum to openly release their collection data is via an application programming interface (API). APIs are a way of structuring data that makes it accessible and transmissible in a machine-readable and dynamic (unlike a static data export to GitHub) way, allowing for communication between software programs. From the perspective of a museum, an API allows users to request data from inside the institution and have it delivered to them in a usable form, freely available with some sort of open licence which is Creative Commons (CC) or CC-like. In making the data public and usable, the museum’s API, like the museum itself, becomes a platform for open, hands-on exploration. This data available via API usually corresponds to the already openly available online catalogue data. Not surprisingly **APIs’ quality depends heavily on catalogues’ data**, including inconsistencies, gaps and thin records. It's not often part of the workflow to assess the data’s quality before it is released openly via API, so it becomes evident how important it is to assess museums’ data, as they’ve been documented and become available via online catalogues.

Although the API technology is by now widely popular in the GLAM world in general, the TaNC audit showed that there is some resistance or delay on the adoption of API from UK GLAM institutions. Almost half (49%) of respondents said they did not have APIs, only 41 respondents (21%) said their institution had ‘an API that allows others to make use of your online collections’ and a further 36 respondents (16%) said that the introduction of an API was pending. Many institutions are dissatisfied with current systems and processes, especially legacy systems without APIs, causing problems integrating with other systems and services. Many institutions were either in the process of, or planning, the move to a new collections management or digital asset management system (DAMS), a state of flux that will last several years (Gosling et al, 2022, 3). Currently, only six of the institutions interviewed for the audit have an Open API to allow third parties to undertake machine-based enquiry or integration with external systems. However, the existing Open APIs have low usage levels and some struggle to perform or stall completely when attempting higher-volume queries for collaborative research projects (Gosling et al, 2022, 5).There is a growing interest aiming to understand how people use cultural heritage institutions’ collection data through APIs: a recent report on evaluating the usability of Museum APIs, from a USA perspecive, focuses on audience’s changing needs, interests, and demographics and offers a number of museum APIs Best Practices in order to support museums’ greater interactivity with their digital

offerings (Villaespesa et al. 2021). Such attempts are of great value as they are offering a way to assess the quality and interest of the available data, to identify specific issues or glitches of the collection data, as well as to evaluate the usability of the current interfaces and available data documentation.

There are also specific API standards developed for cultural heritage content types, such as the International Image Interoperability Framework (IIIF), an open source, community developed standard, able to respond rapidly to needs and desires identified by the GLAM and cultural heritage sectors by providing a set of specifications for interoperable functionality in digital image repositories. Designed by and for the cultural heritage community, the IIIF protocol improves user access to high-resolution digital (images, audio, video, maps and 3D) resources and makes it easier to create rich collection websites with advanced media features. The IIIF protocol provides a world-class user experience to museum, archive and library researchers and (online) visitors by allowing them to to use standard tools (which work with the APIs) to retrieve images from IIIF compatible collections around the world, and to view, share, compare, combine and manipulate images quickly and easily across repositories, in a consistent and shareable way.

The application of IIIF has been assessed as part of the TaNC foundation project "Practical applications of IIIF as a building block towards a digital National Collection" as a way these technical routes and standards might be integrated and adapted to work across collections and mediums for the eventual development of a larger National Collection[5]. Based on their findings, "across the sector, there is some confusion about what 'using IIIF' means, or what benefits it would bring. From those in the IIIF community, there is concern that the lack of IIIF implementation in smaller institutions is contributing to a 'digital divide' in which their collections are under-represented when looking at the GLAM sector, or the future potential for involving institutions in a shared National Collection. There is much to be done to both promote current use of IIIF, and to ensure its adoption, even amongst this set of major institutions. However, many institutions have some variation of open licensing, which would allow the future open implantation of IIIF, and reuse of their digitised image collections" (Padfield et al, 2022, 20).

SMG online catalogue offers the option of a public API which contains the content from the collection website and is detailed at https://www.sciencemuseumgroup.org.uk/about-us/collection/using-our-collection-api.

The API is open to all and a number of innovative ways that external users and organisations are utilising SMG data has been documented below:

- Combining SMG data with external data sets and resources, either individually owned data or other publicly available data.

[5] Padfield, Joseph, Bolland, Charlotte, Fitzgerald, Neil, McLaughlin, Anne, Robson, Glen, & Terras, Melissa. (2022). Practical applications of IIIF as a building block towards a digital National Collection. Zenodo. https://doi.org/10.5281/zenodo.6884885

- Innovative interfaces for viewing and navigating SMG data (Timelines, Force Directed Graphs, Image Recognition).
- Textual analysis of SMG object descriptions and biographies to find patterns and hidden data/terms.
- Physical means of viewing and interacting with SMG collection (Alexa Skills, Motion Sensors).
- Integration with other platforms; such as a TwitterBot that tweets a scientific discovery or pivotal invention each day.
- A display of SMG objects onto a map of your city.

The catalogue data is also available as static data dumps https://github.com/congruence-engine/datasets.

As for the IIIF adaptation from SMG, there is an IIIF manifest available for high-resolution digital resources as part of the online catalogue entry. The IIIF functionality has been recently employed as part of a Getty simulator game, opening up the SMG collection to a new audience, that of the gaming community. However, the lack of documentation around IIIF implementation can be explained by the fact that it is mainly focusing on a community well-aware how to work with it.

3. Linking among collections and records: vocabularies, ontologies, LOD

Although museums have been making their collections' metadata available through APIs for some time now, allowing researchers and users to access and examine closer object records, there has not been the ability to aggregate and link them across multiple museum collections other than in small-scale projects (e.g. American Art Collaborative) or discrete verticals (e.g. ArtUK). The abundance of existing documentation systems and standards in the cultural heritage and, specifically, the museum world, has dictated museum collections and online catalogues to be operating in silos. However, there are a number of techniques mainly related to the Semantic Web, such as controlled vocabularies, ontologies and Linked Open Data technologies, offering ways of interlinking related catalogue records within or across catalogues and institutions.

Controlled vocabularies have been employed not only to support consistency and accuracy in metadata creation but also to improve access and linking among information from a variety of sources, through various knowledge organisation and representation technologies. The links and relationships among the terms in a controlled vocabulary ensure that these connections are defined and maintained, for both cataloguing and retrieval. There are many types of controlled vocabularies, from simple term lists, authority files, to taxonomies, thesauri and complex machine-readable ontologies.
Existing museum collection management systems and catalogues offer the option to incorporate or create controlled vocabularies to further describe and enhance access to their records. Because controlled vocabularies require the use of predefined terms, they can be challenging to adopt and apply. Many of these vocabularies are created and maintained by

research institutions, national and international cultural organizations, and professional societies and associations (e.g. The Getty vocabularies).

Collection catalogue datasets are usually structured in a normalised relational format (usually in a relational database system or RDBMS) with each entity (person, object, place) residing in a separate database TABLE with additional JOIN TABLES holding the relationships between the entity types - in other words, in what are known as information silos. If all datasets were openly published and used the same format for structuring information, it would be possible to link and analyse or interrogate all the datasets at once. Linked Open Data are among the core pillars of the Semantic Web and help break down the information silos that exist between various formats, datasets and collections.
Linked Open Data is a way of publishing structured data using semantic web technologies that allows (meta)data to be enriched, associated and interlinked, so that different representations of the same content can be found. With the implementation of Linked Open Data, alongside other core pillars of the Semantic Web, such as ontologies and graph databases, heritage institutions can connect their collections and records but more importantly to represent data and knowledge in a network of meaningful, flexible and reflexive relations.
Since the early 2000s, there has been an increase in the interest for the design and use of Linked Open Data and other semantic technologies such ontologies and knowledge graph-based systems in humanities computing and in digital cultural heritage environments. However the uptake has been slow, and as a recent literature review conducted as part of the report of the Heritage Connector, an SMG-based TaNC foundation project, barriers to LOD in the cultural heritage sector fall under four broad headings: technical, conceptual, legal and financial Working with LOD at any kind of scale is both time consuming and resource intensive, and human intervention and curation is needed at some point (Winters et al , 2022, 10). Finally, mainly due to the abstract nature and complexity of these technologies and the retrospective application of these techniques to the collection systems, application of LOD in cultural heritage collections has been often held at the **experimentation stage with limited scalability.**

Ontologies are conceptual models resulting from an analysis of a specific knowledge domain, a kind of semantic data model that defines the types of things that exist in our domain and the properties that can be used to describe them. Unlike data models, the fundamental asset of ontologies is their relative independence of particular applications, i.e. an ontology consists of relatively generic knowledge that can be reused by different kinds of applications/tasks. There are a number of standard ontologies for cultural heritage data modelling, like CIDOC Conceptual Reference Model (CIDOC-CRM) or GeoSPARQL. A knowledge graph is made up of interconnected patterns and relationships among a set of individual data points, the nature of which can be continually expanded using a real world ontology: a continually changing informational structure that mediates between a human, the world and a computer.

Among recent developments and applications in the area of Linked Open Data for cultural heritage, the Wikidata ecosystem offers a free to edit knowledge base (aka the semantic database behind Wikipedia), with facts and figures about millions of items as well as mechanisms to interlink among them. Wikidata can be used by those outside the academic or museum community and many of its entities are linked to more generalist knowledge such

as world events or movements. In short, the nature of Wikidata/Wikipedia gives individual entities context. Wikidata has been actively employed by a number of cultural heritage institutions and projects (eg Smithsonian Libraries and Archives, The Digital Archive of Artists' Publishing [DAAP]) as a way to enhance the semantic interoperability and connectivity of their cultural records, but there is still a threshold towards its integration into their business-as-usual workflows and systems.

One of the main technical barriers in the uptake of Linked Open Data technologies in GLAM is that rather than address the entity extraction and disambiguation within the collection management systems themselves, we instead (by and large, outside of some well-funded institutions) are still relying on the linking and disambiguation (to external sources) to happen **outside** the collection systems. In short, the collection systems are continuing to operate in silos, and the task of linking them is seen as an upstream task (possibly as part of a separate digital project) rather than business-as-usual. Incorporating the disambiguation and identification (of entities) to external sources (such as Wikidata) within the collection management systems themselves — rather than as a separate extraction project — would aid future national linking efforts. In a similar vein, the development and wide adaptation of shared vocabularies and keywords across multiple collections, may greatly enhance interoperability and the potential of cross-collection search digitised resources. Both LOD and vocabularies' adaptation is further impeded by the dynamic nature of museum catalogues, making linking often a snapshot of then the links are created: as the catalogue evolves, new or different links might be possible/appropriate, and even some previous links might become in appropriate or need adjusting to reflect the new state of the catalogue.

For the case of SMG, there have been internal taxonomies/thesauri held and structuring information in Mimsy: for railway collections, the Mda Railway Object Name Thesaurus maintained by the Collections Trust is being used. In addition, a legacy import of the Getty ATT thesaurus has been used in-house for structuring catalogue records; the full list of this thesaurus can be viewed as a Categories list (merged across the 5 museums in SMG). SMG has a similar issue with place names, in that we pick values from a legacy Getty TGN import. But none of those place names have the external URI associated with them, which leaves us with no means of disambiguation Birmingham (UK) from Birmingham (USA).

However, there is no consistent use of these thesauri: in many cases there is just a text string 'picked' from Getty ATT rather than an internal pointer to the term in the (structured) thesaurus, prohibiting access to the parent or sibling terms. In addition, there is not the built-in provision to hold any external URLs, UIDs, URIs for SMG terms - it's simply a list of words, so there are no means of linking those terms back to the public version of the thesauri. There is not an (easy means) of identifying with certainty that this value in the SMG catalogue is the same value as this value at Getty. Additionally, the SMG thesauri versions could deviate over time from the main version . Additionally there is no way of using Linked Data technique to find equivalence with that term in other sources such as Wikidata. Which holds an equivalence attribute to the Getty vocab on its record for 'computer'.
While there is currently the technology to re-import and remap Getty ATT into Mimsy, this time with Getty URIs, however finding the time and resources to carry out this work within an overstretched documentation department is challenging.

New terms and content -specific vocabularies have also been developed and used in-house for structuring catalogue records, but they are often not shared publicly or cross-linked to other better-known thesauri, when applicable.

Heritage Connector also proposed a ‘good enough’ record linkage pipeline to Wikidata rather than generating a perfect LOD model, while also creating links within the SMG collection through information retrieval.

4. Discovery and Search: PIDs and generous interfaces

Users of cultural heritage institutions, especially museums, visit online catalogues for different reasons, but their information needs are not always catered for in the way museums present their collections. It is thus crucial for cultural heritage institutions to think more carefully about the information they include when they catalogue their collections, who this information serves, and how the information can be found online. Rather than assuming that online collections information can and will be accessed by everyone the same way, a people-centred design approach to collections search and discovery would focus on the needs of a specific group or groups. Pioneering search and discovery tools based on people-centred design and research have already been introduced in the field, without great uptake, mainly due to lack of available resources, low data quality and limited user-needs research.
To enable identification, consistent access and discovery to museum objects via online catalogues, identifiers or accession numbers are considered a key component in all collection management systems. To fully realise the potential of our national collection, we need identifiers that can bring together collections across institutional boundaries. Persistent Identifiers (PIDs) provide a long-lasting click-able link to a digital object, making content Findable, Accessible, Interoperable and Reusable (FAIR) and enabling data discovery, access, metrics and citation. Supporting wider use of PIDs for collection objects will allow long-term, unambiguous linking of collections across the UK. However, the challenges, utility and wider benefits of PIDs are less well understood across the heritage sector.

Based on the findings of the TaNC foundation project “Persistent Identifiers as IRO Infrastructure”, very few institutions are using third-party, independent services to provide unique, persistent identifiers for catalogue records or digital assets. Most tend to generate identifiers (URLs) based on internal system IDs. These are open to potential change over time which would result in broken links for any dependent resources. The findings of this project highlighted, despite the diversity of the approaches to persistent identifiers that already exist, that a sector-wide approach cannot and should not be overly prescriptive in the types of persistent identifiers that should be used. Individual organisational needs (e.g. linked data metadata approaches vs. collection identification vs. machine readability) and capacity both vary considerably and determine the most appropriate identifier tool(s). But a networked approach can still be built on common principles, functionality and use cases, without necessarily requiring all organisations to use the exact same identifier tools (Kotarski et al, 2022).

Although PIDs would enable better identification, consistent access and discovery to cultural heritage records by bringing together collections across institutional boundaries, in what ways would a large-scale universal generalist collection search interface or (mega)aggregator be better/different than Google Search or Europeana?

However, as is the norm for most cultural collections online, the primary mode of discovery for the content is via search. This limits discovery to those who already know the collection and are looking for something they know is there or speculative searches which may or may not return results depending on what is in the collection, what has been digitised and how it has been catalogued.

Moving beyond structured discovery and search approaches, recently there has been an active interest towards the exploration of online museum collections and catalogues via visual search. In 2015, Mitchel Whitelaw suggested that "[…] search, as the dominant interface to our cultural collections, is inadequate. Keyword search is ungenerous: it demands a query, discourages exploration, and withholds more than it provides. [There is need] instead for *generous interfaces* that better match both the ethos of collecting institutions, and the opportunities of the contemporary web. Generous interfaces provide rich, navigable representations of large digital collections; they invite exploration and support browsing, using overviews to establish context and maintain orientation while revealing detail at multiple scales." (Whitelaw 2015).

By using 'generous interfaces', GLAM institutions are exploring ways to present their collections online in browsable and linkable networks of information that allow users to explore and discover new ideas through meaningful and contextualised relationships. The connectivity of GLAM is facilitated through empowering and user-friendly search technologies such as the **'visual search',** an AI-based method for matching similar images based on their visual characteristics (colour, pattern, shape), rather than a keyword description or search. Visual search, enabled by advanced computer vision (CV) technology, holds huge potential for individuals and organisations and the development of visual search platforms for heritage collections is already underway globally, especially with art museums (for example: ukiyo-e.org, pharosartresearch.org, and Google Arts & Culture initiatives, photography catalogues by keyword, user search by colour or style or other innovative interfaces).

In a similar vein, the Deep Discoveries TaNC foundation project explored the application of computer vision (CV) and explainable artificial intelligence (XAI) methods for enhancing the ability of general audiences and specialist researchers to discover visual collections in new and/or more effective ways. They developed a CV-based search prototype that allowed users to visually articulate their search task, understand how the CV algorithm found similarity between their input image and the returned image results, and to carry out a 'visual dialogue' with the AI to refine their search further. Interestingly, the user-based research underlined that an integrated system should allow users to carry out image retrieval using both visual search and semantic filtering (based on existing or newly generated metadata) in order to provide both discovery-driven and research-specific capabilities (Angelova et al, 2021).

Since the relaunch of their online collections website and the large-scale digitisation program they embarked since 2018, SMG Digital Lab has been actively exploring new forms of discovery for cultural heritage collections online in order to offer new ways of audience engagement and discovery of the wealth of their digitised collection. Through a series of collaborative Collections Remix events and a set of prototypes around generous interfaces and visual search affordances, including Museum in a Tab (Chrome browser extension), Random object generator, What the machine saw, Never been seen, the team has been actively experimenting with various modes of visual search and interface design, while a comprehensive analysis of the various user needs and modes of engagement with the online collection is still missing.[6]

Another key component in this emerging virtual browsing landscape of the SMG collection catalogue is the AI-generated knowledge graphs as a way to facilitate new forms of exploration, discovery and research for digitised cultural heritage collections, as explored by the Heritage Connector TaNC Foundation project, aiming to explore how AI generated knowledge graph from museum collections can facilitate new forms of exploration, discovery and research for digitised cultural heritage collections. Some of the potential benefits we envisioned this knowledge graph could bring were:

- enabling macro-views of the whole collection;
- new and more flexible groupings of items within and across collections;
- richer onward journeys from one collection record, blog post or journal article to the next;
- new entry points from which to begin exploring the collection; and
- new forms of interface, to provide an alternative to keyword search discovery.

Among the interfaces/demos produced are:

1. an interactive streamlit app showing NER and entity linking which uses static data for speed (not hosted at the moment)
2. a bookmarklet to view connections from an SMG collection, blog or journal page
3. a macro visualisation of the whole collection/knowledge graph
4. a visualisation of the combined SMG and V&A collections
5. maps (map 1; map 2) of all the places in the knowledge graph

C. Next steps on museum online catalogues' requirements for a National Collection

This investigation aims to initiate, alongside an online catalogues' assessment framework-matrix, the discussion towards a minimum technical passive provision (see also our what's 'good enough' discussion at the appendix B) for museum online catalogues'

[6] See Appendix A for a user-needs based approach example.

requirements in order to serve as a foundational infrastructure for drawing together a national collection.

Before embarking on developing and applying the technical requirements , the following parameters should be taken into consideration :

- Resource-sensitivity
- Interoperability
- Scalability
- Sustainability
- User-focused
- Efficiency and future-proof of the solution

Level	Description	Functionality	(Top-level) Technical requirements
1	Data only allows read-only **access,** with no provision for being automatically downloaded or retrieved	(Basic) Access (Basic) Search	● Records Digitisation ● Basic data management system (unique ID required) & data import from existing cataloguing entries ● well-formed and discoverable web interface ● easily accessible URLs (not behind logins or access control), ideally includes a unique ID in URL (which can be used to support redirects if URL format or domain has to change)
2	data published as structured data including images and other digital material as separate files	(Basic) Structure (Basic) Access (Basic) Search	● All the above + ● Interoperable, openly accessible, well-documented metadata standard ● Data management system with embedded metadata standard
3	data made accessible in machine-readable form using a non-proprietary format. This data might be stored in Relational Databases (RDB) and could also be accessible in APIs.	Advanced structure Access (Re)use Search	● All the above + ● Public API mechanism (or other publicly accessible & downloadable format/mechanism eg. OAI-PMH, json file, GitHub repo, public data dumps) ● IIIF manifest

4	data made available using a Linked Open Data format (URI), under an open licence allowing them to be connected to central knowledge repositories (eg Wikidata).	Advanced structure Access (Re)use Link Search	• All the above + • LOD mechanism (URIs, RDF) • Structured vocabulary or taxonomy • Field/content-specific ontology
5	connections that an institution makes to a larger context outside of the GLAM context.	Advanced structure Access (Re)use Link Search Advanced Discovery	• All the above + • Linking with external storage repositories such as Wikidata and Getty Research Institute vocabularies which are major data sources in the museum context • Advanced visual search options

Below is an attempt/case-study to apply the above mentioned matrix for the SMG online catalogue, while taking into account comments that have been made through this report:

SMG online catalogue, est 2017

Level	Description	Functionality	(Top-level) Technical requirements
1	Data only allows read-only **access,** with no provision for being automatically downloaded or retrieved	(Basic) Access (Basic) Search	• Records Digitisation The Science Museum Group has published over 300,000 records online. The quantity of digitised objects has also expanded rapidly, from a base of 5% of the collections online with an image, the museum had reached 35% by November 2021. The One Collection project significantly increased the amount of

			content available including c. 200,000 newly digitised objects. ● Basic data management system (unique ID required) and data import from existing cataloguing entries ● well-formed and discoverable web interface https://collection.sciencemuseumgroup.org.uk ● easily accessible URLs (not behind logins or access control), ideally includes a unique ID in URL (which can be used to support redirects if URL format or domain has to change) https://collection.sciencemuseumgroup.org.uk/objects/**co34900**/h-t-w-t-unit-100-watt-no-416-h-t-w-t-unit
2	data published as structured data including images and other digital material as separate files	(Basic) Structure (Basic) Access (Basic) Search	● All the above + ● Interoperable, openly accessible, well-documented metadata standard Collection Records conform to Spectrum standards, with additional rules internal to cover cataloguing needs not covered by Spectrum. **Openly accessible:** Yes, via a simple to use, public facing and open JSON API. IIIF image delivery for images. Metadata is CC0, descriptions CC-BY-NC 4.0, most images also CC-BY-NC-SA 4.0 where rights owned by SMG. **Interoperable:** Limited interoperability as our API does not **currently** contain external identifiers (either Wikidata or Getty) for object names, geographical places or materials

			due to a lack of those external identifiers within SMG's internal cataloguing systems. However, as part of SMG's Collection Online 2.0(CO2.0) project (2023/2024) it will contain Wikidata identifiers identified as part of the Heritage Connectors project, for a limited subset of people and companies. **Well-documented metadata standard:** **(Data) Yes**, in terms of JSON being an easy to parse format. **No**, in terms of a formal standard like CIDOC-CRM. Worth noting / mentioning that as far as we are aware, this hasn't hindered anyone using our Collection data, despite running multiple 'Hack Days' where users have used our API with no advanced training. **(Images)** Publicly accessible IIIF presentation manifests and IIIF image delivery endpoint. ● Data management system with embedded metadata standard Three (3) SMG's Information Systems are currently employed: Mimsy is the collection management system used for objects, whereas Adlib is used for archival collections and Koha for library collections.
3	data made accessible in machine-readable form using a non-proprietary format. This data might be stored in Relational Databases	Advanced structure Access (Re)use	● All the above + ● Public API mechanism (or other publicly accessible & downloadable format/mechanism eg. OAI-PMH, JSON file, GitHub repo, public data dumps) SMG online catalogue offers the option of a public API which contains

	(RDB) and could also be accessible in APIs.	Search	the content from the collection website and is detailed at https://www.sciencemuseumgroup.org.uk/about-us/collection/using-our-collection-api. • IIIF manifest There is an IIIF manifest available for high-resolution digital resources as part of the online catalogue entry. However, the lack of documentation around IIIF implementation can be explained by the fact that it is mainly focusing on a community well-aware how to work with it.
4	data made available using a Linked Open Data format (URI), under an open licence allowing them to be connected to central knowledge repositories (eg Wikidata).	Advanced structure Access (Re)use Link Search	• All the above + • LOD mechanism (URIs, RDF) • Structured vocabulary or taxonomy • Field/content-specific ontology There have been internal taxonomies/thesauri held and structuring information in Mimsy: for railway collections, the Mda Railway Object Name Thesaurus maintained by the Collections Trust is being used. In addition, a legacy import of the Getty ATT thesaurus has been used in-house for structuring catalogue records; the full list of this thesaurus can be viewed as a Categories list (merged across the 5 museums in SMG). SMG has a similar issue with place names, in that we pick values from a legacy Getty TGN import. But none of those place names have the external URI associated with them, which leaves us with no means of disambiguation .

5	connections that an institution makes to a larger context outside of the GLAM context.	Advanced structure Access (Re)use Link Search Advanced Discovery	• All the above + • Linking with external storage repositories such as Wikidata and Getty Research Institute vocabularies which are major data sources in the museum context Heritage Connector developed a customised OS record linkage pipeline with Wikidata, allowing SMG records to be linked via Information Retrieval to the knowledge graph or /and Wikidata. The Collection API **does not currently** include Getty UIDs. The ability to achieve this is dependent on backend work to cataloguing standards / systems, which do not currently hold those UID values. As part of SMG's CO2.0 project (2023/2024) the SMG CO API will contain Wikidata identifiers that were disambiguated as part of the Heritage Connectors project, for a limited subset of people and companies. *(These Wikidata values are already in Mimsy, however as part of the CO2.0 project they will be pushed through to the public API as well)* • Advanced visual search options Here are a couple of visual search prototypes currently available for the SMG online collection: Chrome Extension, Random object generator, What the machine saw, Never been seen.

Overall comments:

The SMG online collection catalogue and dataset score a good **4+**, with **5** TBC as the requirements are currently in an experimental/development stage. What the annotated matrix shows is that, although the SMG online catalogue already succeeds to meet to a good standard the necessary technical requirements that will enable it to be linked with other collections and resources, there are areas of improvement where focused action is required.

Appendix

A. Why? Towards a user-needs based approach on requirements' planning

Why link? Why particular standards or vocabularies? Why build APIs? To what end? What end uses does this work support?

In the museum / tech sector we often forget to ask this question when having these conversations. But in many ways, it's the most important question of all, as it helps lead to sensible and pragmatic solutions that solve real users' needs. Through this investigation we want to put forward the idea of starting with and thinking about user needs and then asking, "what is needed to support those particular needs". Such a user-needs focused approach often leads to different (better) answers and more pragmatic and useful solutions than just thinking in the abstract.

—

Below are a couple of use cases/scenarios for online cultural heritage collections and catalogues that would help us prioritise our work towards requirements' planning:

1. New research questions can be asked
    a. Visualisations and analysis to understand the collections at scale
    b. Cross-collection analysis
    c. Data mining
2. (Prohibitively) Time consuming tasks can be approached
3. Greater accessibility and likely to a wider audience
4. Creative/Artistic projects undertaken
5. Bespoke/custom interfaces/projects which work for particular audiences ("all the nation's wallpaper collections searchable by colour and pattern" etc.)
6. Experimental projects

1/ General Discovery
Use case:

- Support general users research needs
- Homework
- Local interest groups
- General academic/scholarly research

How:

- Appearing in Google and general web searches.
- Allow content to be harvested by LLMS (e.g. ChatGPT) and search engine web crawlers (e.g. Google's) to increase discoverability

Needs:

- Specific formats and APIs not required, well formed and discoverable HTML important
- Long lived (10+ year plan), easily accessible URLs (not behind logins or access control), ideally includes a unique ID in URL (which can be used to support redirects if URL format or domain has to change)
- Adding popular / widely used external identifiers (Wikidata/Getty/Library of Congress Subjects etc. (to People (John Smith vs John Smith), Place (Birmingham vs Birmingham) and Term/ObjectName/Subject pages where applicable helps uniquely identify these pages ie. Plate (camera) vs Plate (railway) vs Plate (ceramics). Can just be a HTML link on page or https://json-ld.org tag in HTML.

2/ Data focused research

Use case:

- Support users who wish to evaluate museum collections as **data**, such as collecting trends over time by material or object type by county, automated timelines, geographical mapping, collecting dining menus or newspapers from multiple sources to analyse public trends or habits.

How:

- Provide access to data in machine readable format
- Level of specificity will depend on users needs, advanced use cases may be better handled manually as requests occur, adding functionality back to the general platform if the same requests occur frequently.
- Support common use cases via API and public data dumps. Tread carefully in building an over complex API that then becomes a technical barrier to many. Do not underestimate the need to liaise and support users with more advanced research needs.

Needs:

- Easy to use API or exports of datasets
- Disambiguation of key entity types
- May need to support some research with custom exports and personal discussions.

3/ Aggregation sites
Use cases:

- Vertical aggregators which have specific search interfaces for particular types of content. ie ArtUK (UK Art), Archeology (Un-pathed Waters), WWI (A Street Near You)

How:

- Provide APIs or regular data dumps that expose data including and specific terms / entities used by Vertical aggregator, ideally disambiguated (via an external URI / UID)

Needs:

- Provide APIs or regular data dumps that expose data including and specific terms / entities used by Vertical aggregator, ideally disambiguated (via an external URI / UID)

Issues/Questions:

- Is the third-party site offering something unique above your own site or Google in terms of a specific interface into particular content. Is the aggregation adding real value.
- Is your data relatively static, in which case a data export may suffice or do you need a means of providing regular updates
- Who does the data alignment / disambiguation or terms or entities, partially with regards to a specific subject area or domain that the aggregator specialises in? Can you bring back any external disambiguation work back into your collection system?

4/ Ad-Hoc / specialised research projects
Use cases:

- Small scale research on specific areas / records

How:

- May be better handled by museum staff as required information may not have been digitised or electronically catalogued.
- Needs may be too specific for generalised API or data dumps.

Needs:

- Provide museum staff with ability to assist, discover, access and digitise materials on demand
- Setup a process for relevant material to be added back to digital catalogue where appropriate

B. What's 'good-enough' for a national collection ?

One of the biggest challenges the Congruence Engine - and the TaNC programme at-large - is facing is how to define the technical requirements of a national-level infrastructure not by design but in a 'business-as-usual' approach: by proposing ways to enable connections among cultural institutions with existing collections, systems, workflows and communities already in place. In order to better address this challenge, we propose to adopt a 'good enough' approach in our technical endeavours and recommendations.'Good enough' is also an approach that might be useful to employ when dealing with experimental techniques requiring a lot of customisation, specialised skills and resources in order to be applied in a large-scale framework.

‘Good enough’ scenarios:

- To what extent a LOD approach needs to be perfect (X explicitly made Y) and to what extent it can be a bit vague (X is related in some way to Y)?

  The first is obviously better, but at scale only works if everything is described that way which is hard to test and hard to do. Whereas the second one potentially is easier to scale and immediately adds value to users and could be built into user interfaces more rapidly, but would leave some of the unravelling of what's going on to users. Arguably, the second one would help with exploration and discovery whereas the first one could help solve particular research questions at scale.

‘Good enough’ solutions:

- lightweight add-on software that can be integrated alongside existing systems
- shared infrastructure approaches in support of smaller institutions with little or no technical capacity
- enrich and/or update Wikidata entries

Bibliography

Angelova, Lora, Ogden, Bernard, Craig, Jack, Chandrapal, Hari, & Manandhar, Dipu. (2021). Deep Discoveries: A Towards a National Collection Foundation Project Final Report. Zenodo. https://doi.org/10.5281/zenodo.5710412

Gosling, Kevin, McKenna, Gordon, & Cooper, Adrian. (2022). Digital Collections Audit. Zenodo. https://doi.org/10.5281/zenodo.6379581

Kotarski, Rachael, Kirby, Jack, Madden, Frances, Mitchell, Lorna, Padfield, Joseph, Page, Roderic, Palmer, Richard, & Woodburn, Matt. (2022). Persistent Identifiers as IRO Infrastructure: A Towards a National Collection Foundation Project Final Report. Zenodo. https://doi.org/10.5281/zenodo.6359926

Anne Luther (2022), Digitization And Data Management In Museums, 23/10/2022

Padfield, Joseph, Bolland, Charlotte, Fitzgerald, Neil, McLaughlin, Anne, Robson, Glen, & Terras, Melissa. (2022). Practical applications of IIIF as a building block towards a digital National Collection. Zenodo. https://doi.org/10.5281/zenodo.6884885

Padilla, Thomas, Allen, Laurie, Frost, Hannah, Potvin, Sarah, Russey Roke, Elizabeth, & Varner, Stewart. (2019). Final Report --- Always Already Computational: Collections as Data (Version 1). Zenodo. https://doi.org/10.5281/zenodo.3152935

Villaespesa, E., Nadel, K., Estigarribia, A., Tankha, M. and Korshakova, E. (2021), Evaluating the Usability of Museum APIs Report , https://prattdx.org/wp-content/uploads/2021/04/Report-Evaluating-the-Usability-of-Museum-APIs.pdf

Mitchell Whitelaw (2015), “Generous Interfaces for Digital Cultural Collections”, *Digital Humanities Quarterly*, Volume 9 Number 1
http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html

Winters, Jane, Stack, John, Dutia, Kalyan, Unwin, Jamie, Lewis, Rhiannon, Palmer, Richard, & Wolff, Angela. (2022). Heritage Connector: A Towards a National Collection Foundation Project Final Report. Zenodo. https://doi.org/10.5281/zenodo.6022678