Information Research - Vol. 30 No. CoLIS (2025)

Exploring the adoption of collections as data in the GLAM context

Meltem Dişli, Nele Gabriëls, Sally Chambers, Sarah Ames, Beth Knazook, Gustavo Candela

DOI: https://doi.org/10.47989/ir30CoLIS52252

Abstract

Introduction. This research investigates the adoption of *collections as data* in GLAM institutions, aiming to highlight best practices for computational use and machine-readable sharing of digital collections.

Method. Drawing on the existing literature regarding collections as data, this research identifies key aspects that should be considered and assesses selected datasets based on these aspects.

Analysis. The selected datasets were analysed based on key aspects to identify best practices and gaps.

Results. The findings revealed areas needing improvement, including impact, ethical rights, citation and limitation information.

Conclusion. The work has identified key aspects and provided examples of best practices. Future work includes the inclusion of additional aspects to analyse the publication of collections as data such as sustainability and discoverability.

Introduction

GLAM (Galleries, Libraries, Archives and Museums) have been exploring new ways to make their digital – both digitised and born-digital – collections available to foster responsible reuse. New initiatives such as GLAM Labs (https://glamlabs.io/) and collections as data (https://collectionsasdata.github.io/), in combination with the FAIR (https://www.go-fair.org/fair-principles/) and CARE (https://www.gida-global.org/care) principles, have recently emerged to promote computational and responsible use of digital collections including techniques based on the use of artificial intelligence (AI) and machine learning (Wilkinson et al., 2016; Padilla et al., 2019; Carroll et al., 2020; Alkemade et al., 2023; Candela, Gabriëls, et al., 2023; Lee, 2023). Specific working groups have been created in the context of data organisations such as the Research Data Alliance Interest Group on Collections as Data (Research Data Alliance, 2024). Other initiatives, such as AI4LAM, are focused on the application of AI to cultural heritage content (AI4LAM, 2024).

Previous work has tackled the analysis of how digital collections can be made available for computational use (Padilla et al., 2019; Candela, Chambers, and Irollo, 2023; Candela, Sáez, et al., 2023; Dişli, 2024) and encouraged global consensus-building on the ethical responsibility and perceived benefits for GLAM professionals in presenting collections as open data resources (Knazook et al., 2023; Padilla et al., 2023). The copyright and licensing conditions for collections as data have also been discussed (Dişli and Candela, 2025), and remain an ongoing challenge for the release of GLAM data collections. Despite all these efforts, there is still room for improvement in the establishment of best practices and standardisation. The testing of new methods and approaches developed very recently has been undertaken mainly by well-funded, national organisations, primarily in the Global North.

In cases where organisations have made significant efforts to adopt *collections as data* practices, several challenges have emerged, including budget and resourcing constraints, technical limitations, required skills, qualified personnel, traditional and established trends and copyright issues (Candela et al., 2022; Dişli and Tonta, 2023). Overcoming these challenges, addressing the encountered gaps, identifying potential problems and establishing standardised processes are crucial for assessing current initiatives.

The purpose of this work is to provide an analysis of the current state of the adoption of collections as data principles in GLAM institutions. In line with this purpose, selected GLAM datasets published in a machine-actionable format are analysed and assessed based on specific criteria. This work also aims to contribute to the adoption of best practices and the development of the necessary standards and guidelines for the publication of machine-actionable collections.

This work seeks to answer the following research questions: 1) to what extent do existing datasets published in alignment with the collections as data approach meet the identified aspects?; 2) what are examples of best practices for promoting the broader adoption of collections as data?; 3) what are the key challenges faced by GLAM institutions in the adoption of collections as data according to best practices and guidelines; and 4) what actions are needed to address these challenges?

The main contribution of this work is to analyse evidence of computationally ready collections as a way of evaluating a shift in practice. The results will contribute to identifying best practices for publishing machine-actionable collections in the GLAM domain and to establishing standards in this regard.

Related work

GLAM institutions have been exploring new approaches for enhancing the accessibility and reusability of their collections and metadata. Notable initiatives such as Collections as Data, the International GLAM Labs Community and the Research Data Alliance Interest Group on Collections

as Data (https://www.rd-alliance.org/groups/collections-as-data-ig/posts/?post=170960) promote the publication of machine-actionable collections. Other domains such as web archiving, which has traditionally been carried out by GLAM organisations, have also explored the publication of machine-actionable collections and its documentation (Candela, Chambers, and Sherratt, 2023; Carlos Lelkes-Rarugal, 2024).

At the core of collections as data efforts lies the longstanding work on making cultural heritage collections available openly. In this context, various projects and policies have been developed over the years. To support the open release of cultural heritage collections, the Open GLAM (https://openglam.org/) movement was launched, accompanied by the development of a Declaration on the subject (Scann et al., 2020). Additionally, Creative Commons has contributed to open access efforts in the cultural heritage domain (Creative Commons, 2025). Platforms such as Europeana also actively support and encourage open access (Europeana Pro, 2025a).

Over the past few years, institutions such as the Getty Research Institute (https://www.getty.edu/research/tools/vocabularies/lod/index.html) and the Rijksmuseum Data Services (https://data.rijksmuseum.nl/) have been exploring new ways to publish digital collections. Previous works have focused on various methods to make GLAM content available in the form of machine-actionable collections. Some examples include: i) the application of the FAIR principles to libraries, archives and museums (Koster and Woutersen-Windhouwer, 2018); ii) a checklist providing relevant items to consider when publishing collections as data (Candela, Gabriëls, et al., 2023); iii) a datasheet approach to provide clear, transparent and standardised documentation for digital collections (Alkemade et al., 2023); iv) the identification of key aspects (e.g., formats, licence or terms of use) to consider when publishing machine-actionable collections (Candela, Sáez, et al., 2023); v) a workflow to publish collections as data (Candela, Chambers and Irollo, 2023); and vi) a conceptual model for cultural heritage collections as data (Dişli, 2024).

In parallel, there has been a growing interest in data spaces as new infrastructures with which to share data in a secure and trustworthy environment (Dobreva et al., 2022). In this context, a workflow was recently made available to facilitate the publication of datasets (https://marketplace.sshopencloud.eu/workflow/I3JvP6). In the UK context, a recent report provides detailed training materials, along with case study examples, to support cultural heritage institutions and their funders in the building of a unified UK digital collection to unlock their full potential (Bailey et al., 2024), as well as the exploration of infrastructural futures for digital cultural heritage (Ross, et al., 2024). In parallel, and given the increase of data published in the last years, GLAM institutions have started to restrict access to collections for training commercial AI (https://www.kb.nl/en/news/kb-restricts-access-to-collections-for-training-commercial-ai) as well as to promote the *responsible adoption* and use of AI (National Library of Scotland, 2024).

When making digital collections available, publishing examples of use is essential to show researchers how the data can be reused. In this sense, Jupyter Notebooks have emerged as a reproducible tool to provide code and documentation. Many institutions have started to employ Jupyter Notebooks to introduce researchers to their digital collections (Candela, Chambers and Sherratt, 2023; Maemura and Byrne, 2024).

Assessing reuse of digital collections is crucial for the institutions in order to demonstrate their impact (Hughes, 2012; Tanner, 2016; Warwick and Bailey-Ross, 2020). However, this can be a challenging task for GLAM institutions. Toolkits such as D-CRAFT (https://reuse.diglib.org/) provide different methods to assess the reuse and impact of digital collections as well as tutorials and examples of use. These approaches are essential for GLAM institutions and projects in need of measuring the impact of their work.

Despite all these efforts, to the best of our knowledge, there is no analysis of existing collections published as datasets in terms of criteria based on the combination of FAIR, CARE and impact

aspects. This can be useful for GLAM institutions willing to adopt the collections as data principles following best practices and guidelines.

Methodology

A literature review of existing approaches to publish machine-actionable collections described in this work (see section Related work) was performed to identify relevant and key items to consider when making available machine-actionable collections. This work is mainly based on previous work published by relevant communities and initiatives relating to the adoption and publication of digital collections suitable for computational use.

Building on this existing work, relevant topics were selected such as documentation covering various features like curation rationale, impact on the community and the institution, usage details and publication methods. This research is based on the FAIR and CARE principles. While most of the selected aspects align with the FAIR principles, particularly reusability and interoperability principles, certain aspects also emphasise bias and rights, which are central to the CARE principles. In addition to these principles, some examples of relevant developments used to identify key aspects to publish collections as data include: i) the checklist and workflow to publish collections as data since it provides recommendations and best practices focused on small and medium-sized institutions, and data spaces for cultural heritage (Candela, Gabriëls, et al., 2023, Candela, Chambers and Irollo, 2023); ii) the datasheets for cultural heritage datasets concerning the provision of documentation based on key dimensions such as data provenance and biases (Alkemade et al., 2023); and iii) the D-CRAFT tool providing means to measure impact and reuse of digital collections (D-CRAFT, 2025). As a result, sixteen items were selected as is shown in Table 1.

No	Aspects	Descriptions
1	Documentation	Documentation about curation rationale, provenance and source data.
2	Examples of use	Practical applications or research examples showing the use of the dataset.
3	Accuracy	Accuracy of data, based on OCR quality, metadata quality e.g., LOD.
4	Bias	Discussion of biases (See for example, the DE-BIAS project: https://pro.europeana.eu/project/de-bias).
5	Limitations	Known limitations e.g., missing information.
6	Impact	Information regarding the impact and reuse of the digital content.
7	Terms of use	Conditions and restrictions governing the access and usage of the dataset, including licence, rights and limitations on certain use cases.
8	Licence	The terms under which the dataset can be used, shared or modified (e.g. Creative Commons (https://creativecommons.org/) and Traditional Knowledge Labels (https://localcontexts.org/labels/traditional-knowledge-labels/)).
9	Rights	Legal ownership of the data. Beyond the licence, this may include moral or ethical rights, such as community rights.
10	Citation information	Citation guidelines to help users properly cite the dataset.
11	Identifier	A unique identifier for datasets, such as a DOI (Digital Object Identifier).
12	Collaborative edition platforms	Inclusion of the dataset in collaborative edition environments such as the Social Sciences and Humanities Open Marketplace and Wikidata.
13	Metadata standards	Metadata structured in accordance with international standards such as MARC and Dublin Core.
14	Machine-readable metadata	The structuring of metadata in a way that allows machines to read and interpret it.
15	Machine-readable file format	File formats that are structured and formatted for automated processing (e.g., JSON, XML, RDF, CSV).
16	API	Providing an Application Programming Interface to access the dataset. This assessment considers the API facilitates access to the dataset itself.

Table 1. Key aspects to consider when making available collections as data.

In Table 1, each aspect has been assigned a number, which is used below in the presentation of findings for brevity.

To address the research questions, machine-actionable datasets from various GLAM institutions were selected through extensive web-based research. The datasets are all published under open licences enabling reuse and included on Labs websites or GLAM research repositories. The selection process prioritised datasets from diverse institutions and data types (including metadata, text or multimedia content) to ensure a broad representation of different approaches. Additionally, datasets from institutions recognised as best practice examples in providing collections as data were included. These datasets were also chosen as examples because they were observed to be published in accordance with the FAIR and CARE principles.

The datasets were assessed using the document analysis method, a systematic approach for reviewing and evaluating printed or electronic materials (Bowen, 2009), focusing on aspects such as metadata information, formats and licensing details. In reporting the findings, a ✓ symbol in Table 2 indicates that a dataset addressed the aspect, while an ✗ denotes its absence. The analysis aimed to identify the strengths and weaknesses of the datasets, determine the extent to which the identified aspects are fulfilled, and provide an overall evaluation of their machine-actionability. The results are expected to serve as examples of best practices for GLAM institutions interested in adopting the collections as data approach.

Results

The method described in the previous section was applied to a selection of datasets suitable for computational use and made available by GLAM institutions. The datasets mostly consist of bibliographic data, OCR texts, images and metadata, including information on artworks, manuscripts, archaeological finds and web archives from GLAM institutions. While the majority of the datasets seem to address more than half of the criteria, five of them fail to address most of the criteria.

Dataset	Type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
London Terrorist Attack 7th July 2005 - Collection Seed List	Meta-data	✓	✗	✓	✗	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗	✓	✗	9
Alter Realkatalog of Berlin State Library	Meta-data	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	✓	✓	✗	12
State Library of Queensland - Photographs 1914-1918	Images	✓	✗	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗	5
Biblioteca Virtual Miguel de Cervantes	Bibliog-raphic data	✓	✓	✓	✗	✗	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	11
Timarit.is	OCR and images	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	7
Chronicling America	OCR and images	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	11
Austrian National Library - Historical Postcards	Images and Meta-data	✓	✓	✓	✗	✗	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓	✓	8
Rijksmuseum	Bibliog-raphic data	✓	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	8
National Library Diet, Japan	Bibliog-raphic data	✓	✓	✓	✗	✗	✓	✓	✓	✗	✓	✓	✗	✗	✗	✓	✓	10
Smithsonian American Art Museum as LOD	Meta-data	✓	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✓	✓	✓	✓	7
CoinSampo - Finland	Meta-data	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✓	✓	✓	7

OPenn - Collection of Indic Manuscripts	Images and Meta-data	✓	✓	✗	✗	✗	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	**6**
KU Leuven Libraries Magister Dixit Collection	Meta-data	✓	✓	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	✗	**9**
Total		**13**	**10**	**7**	**1**	**2**	**4**	**9**	**11**	**1**	**6**	**5**	**4**	**9**	**8**	**12**	**8**	

Table 2. Assessment results according to the digital collections selected.

Most datasets are satisfactory in terms of documentation, examples of use, accuracy, terms of use, licence, metadata standards, machine-readable metadata and file formats and API. The documentation includes information about the dataset, provenance, curation rationale and source data. The '*examples of use*' criterion was based on providing descriptions of potential uses and existing tools or prototypes. Accuracy has been assessed based on factors such as whether the texts are OCR-processed, the clarity of images, and the completeness and accuracy of metadata. It has also been examined whether information about these aspects is included in the dataset.

Despite satisfactory aspects, the dataset analysis shows that improvements are needed in several aspects. For instance, with the exception of one dataset, none discuss bias and only two datasets provide information on limitations. Additionally, apart from four datasets, no information is provided regarding the reuse and impact of the datasets. Only one dataset includes information about moral and ethical rights. Guidance on how to cite datasets is not widely adopted and only a few datasets use unique identifiers, such as DOIs. The inclusion of datasets in collaborative platforms, such as Wikidata, also has not been widely embraced.

Discussion

This work assesses the adoption of collections as data in GLAM institutions based on certain criteria derived from the FAIR and CARE principles, aiming to answer four key research questions outlined in the Introduction section.

Regarding the first research question, the results show that while datasets largely comply with FAIR principles, further work is needed to align them with CARE principles. Related to the second question, exemplary datasets have been identified. In particular, datasets that address most of the aspects are good examples of adopting collections as data. However, it should be noted that all the datasets have aspects that require improvement.

Concerning the third research question, and considering previous works (Candela et al., 2022; Dişli and Tonta, 2023), this research identified several challenges in the adoption of certain aspects of collections as data. In identifying these challenges, the FAIR and CARE principles have also been taken into consideration. The lack of information on moral and ethical rights makes it difficult to publish these datasets in alignment with the CARE principles, as does the limited information about bias in the selection of dataset content. Furthermore, for some datasets in this selection, the absence of unique identifiers, the exclusion of the dataset from collaborative editing platforms and the lack of citation information make it difficult to ensure their findability, accessibility and reusability. Additionally, the impact values of the datasets were found to be low, suggesting that they are either not sufficiently promoted or not provided in formats conducive to reuse. All these

issues can be due to several reasons related to a lack of funding and resources. However, GLAM institutions are starting to consider aspects such as bias to provide documentation about their datasets (Europeana Pro, 2025b).

In response to the fourth research question, several recommendations have been proposed for GLAM institutions wishing to adopt collections as data. Institutions should consider the challenges identified above in order to strengthen the FAIR and CARE nature of their datasets. To this end, the findings of this study serve as a valuable reference as it highlights exemplary practices that offer solutions to the challenges encountered while at the same time identifying some of these exemplary practices' shortcomings. In creating and improving their machine-actionable collections, institutions can build upon these existing examples by adopting their strengths while also being aware of and addressing their limitations. This practice will contribute to a more comprehensive and advanced adoption of the collections as data principles.

Conclusions and future work

This work presents an analysis of current trends and best practices to make available digital collections suitable for computational use. Several initiatives promote and foster the adoption of the collections as data principles in the GLAM domain. The results show a wide diversity of approaches when making available digital collections. They also show that there is evidence of ease of alignment between emerging and historical practices for the publication of digital collections, which can be used to encourage institutions to adopt new methods when making available their digital collections.

Though efforts were made to select datasets from diverse institutions and with various types of content such as bibliographic data, OCR texts, images and metadata, the sample remains a limited selection. The present study can be extended by including additional datasets with a higher level of variety regarding features such as language, content and country of origin. Future work to be explored also includes investigating additional aspects such as sustainability, archiving and digital preservation of collections as data. Additionally, the aggregation of machine-actionable collections on shared platforms through protocols like OAI-PMH (https://www.openarchives.org/pmh/) and emergent trends based on the use of data spaces, as well as interoperability issues concerning the use of vocabularies to describe datasets such as DCAT (https://www.w3.org/TR/vocab-dcat-3/), are aspects that need to be considered.

About the authors

Meltem Dişli is a research assistant at Hacettepe University, Department of Information Management. She completed her PhD in the same department, focusing on collections as data. Her research interests are collections as data, computational access, cultural heritage and digital humanities. She can be contacted at meltem.disli@hacettepe.edu.tr

Nele Gabriëls (PhD, Historical Musicology) is responsible for digitisation at KU Leuven Libraries. Working on digitisation process development and policy preparation, she focuses on readying the library's physical collections for computational use in a GLAM Lab environment. Her current research focuses on collections as data. She can be contacted at nele.gabriels@kuleuven.be

Sally Chambers is Head of Research Infrastructures Services, The British Library and on the DARIAH Board of Directors. Previously she was DATA-KBR-BE project coordinator facilitating data-level access to KBR's digital collections and Digital Humanities Research Coordinator at the Ghent Centre for Digital Humanities. She can be contacted at sally.chambers@bl.uk

Sarah Ames is Digital Scholarship Librarian at the National Library of Scotland, responsible for Digital Research and developing the library's open data platform, Data Foundry. She collaborates

on a wide range of research initiatives bringing together digital tools and methods to explore the national collections. She can be contacted at sarah.ames@nls.uk

Beth Knazook is Research Data Project Manager at the Digital Repository of Ireland. For her PhD in Art and Visual Culture from Western University in Canada her focus was on the history of photographically illustrated books in Canada and sharing humanities research collections as data. She can be contacted at b.knazook@ria.ie

Gustavo Candela is a lecturer in Computer Science at the University of Alicante. He holds a PhD in Computer Science from the University of Alicante. His research interests are the semantic web, collections as data, reproducibility and data spaces in GLAM institutions. He can be contacted at gcandela@ua.es

References

AI4LAM. (2024). *AI for Libraries, Archives, and Museums*. https://sites.google.com/view/ai4lam (Archived by the Internet Archive at https://web.archive.org/web/20250220060129/https://sites.google.com/view/ai4lam)

Alkemade, H., Claeyssens, S., Colavizza, G., Freire, N., Lehmann, J., Neudecker, C., Osti, G., & Strien, D. van. (2023). Datasheets for digital cultural heritage datasets (1). *Journal of Digital Humanities Data*, 9(1), Article 1. https://doi.org/10.5334/johd.124

Bailey, R., Pereda, J., Michaels, C., & Callahan, T. (2024). *Unlocking the potential of digital collections. A call to action. Towards a national collection*. Arts and Humanities Research Council. https://doi.org/10.5281/zenodo.13838916

Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. https://doi.org/10.3316/QRJ0902027

Candela, G., Chambers, S., & Irollo, A. (2023). A workflow to publish Collections as Data: the case of Cultural Heritage data spaces. *Social Sciences & Humanities Open Marketplace*. https://marketplace.sshopencloud.eu/workflow/I3JvP6

Candela, G., Chambers, S., & Sherratt, T. (2023). An approach to assess the quality of Jupyter projects published by GLAM institutions. *Journal of the Association for Information Science and Technology*, 74(13), 1550–1564. https://doi.org/10.1002/asi.24835

Candela, G., Gabriëls, N., Chambers, S., Pham, T.-A., Ames, S., Fitzgerald, N., Hofmann, K., Harbo, V., Potter, A., Ferriter, M., Manchester, E., Irollo, A., Van Keer, E., Mahey, M., Holownia, O., & Dobreva, M. (2023). A checklist to publish collections as data in GLAM institutions. *Global Knowledge, Memory and Communication*. https://doi.org/10.1108/GKMC-06-2023-0195

Candela, G., Sáez, D., Escobar, P., & Marco-Such, M. (2022). Reusing digital collections from GLAM institutions. *Journal of Information Science*, 48(2), 251–267. https://doi.org/10.1177/0165551520950246

Candela, G., Sáez, M.-D., Escobar, P., & Marco-Such, M. (2023). A benchmark of Spanish language datasets for computationally driven research. *Journal of Information Science*, 49(6), 1451–1461. https://doi.org/10.1177/01655515211060530

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., &

Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43. https://doi.org/10.5334/dsj-2020-043

Creative Commons. (2025). *Open culture: open access to cultural heritage*. In Creative Commons Certificate for Educators, Academic Librarians, and Open Culture. https://certificates.creativecommons.org/cccertedu/chapter/5-1-open-glam-open-access-to-cultural-heritage/ (Archived by the Internet Archive at https://web.archive.org/web/20250326073723/https://certificates.creativecommons.org/cccertedu/chapter/5-1-open-glam-open-access-to-cultural-heritage/)

D-CRAFT (2025). *Measure the impact of your digital collections*. https://reuse.diglib.org/ (Archived by the Internet Archive at https://web.archive.org/web/20250214132129/https://reuse.diglib.org/)

Dişli, M. (2024). Veri Olarak Kültürel Miras Koleksiyonları [Cultural Heritage Collections as Data]. Doctoral Dissertation, Hacettepe University. https://openaccess.hacettepe.edu.tr/xmlui/handle/11655/34990

Dişli, M., & Candela, G. (2025). Copyright and licencing for cultural heritage collections as data. *Journal of Open Humanities Data*, 11, 1–14. https://doi. org/10.5334/johd.263

Dişli, M., & Tonta, Y. (2023). Veri Olarak Kültürel Miras Koleksiyonları [Cultural Heritage Collections as Data]. *Türk Kütüphaneciliği*, 37(3), 191-214. https://doi.org/10.24146/tk.1317445

Dobreva, M., Stefanov, K., & Ivanova, K. (2022). Data spaces for cultural heritage: insights from GLAM Innovation Labs. In Proceedings of *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries: 24th International Conference on Asian Digital Libraries*, ICADL Hanoi, Vietnam, 30 November – 2 December 2022 (pp. 492–500). https://doi.org/10.1007/978-3-031-21756-2_41

Europeana Pro. (2025a). *Open and reusable digital cultural heritage*. https://pro.europeana.eu/page/open-and-reusable-digital-cultural-heritage (Archived by the Internet Archive at https://web/20250326074833/https://pro.europeana.eu/page/open-and-reusable-digital-cultural-heritage)

Europeana Pro. (2025b). *DE-BIAS - Detecting and cur(at)ing harmful language in cultural heritage collections*. https://pro.europeana.eu/project/de-bias (Archived by the Internet Archive at https://web.archive.org/web/20250221101643/https://pro.europeana.eu/project/de-bias)

Hughes, L. M. (Ed.). (2012). *Evaluating and measuring the value, use and impact of digital collections*. Facet Publishing.

Koster, L. and Woutersen-Windhouwer, S. (2018). FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections. *Code{4}lib Journal*, 40. https://journal.code4lib.org/articles/13427 (Archived by the Internet Archive at https://web.archive.org/web/20250326091208/https://journal.code4lib.org/articles/13427)

Knazook, B., Murphy, J., Barner, K., Cassidy, K., Claeyssens, S., Cortese, C., Manchester, E. J., Padilla, T., Reijerkerk, D., Robson, G., Schmidt, A., Sherratt, T., & Warren, M. (2023). WorldFAIR Project (D13.2) *Cultural heritage image sharing recommendations report* (Version 1). https://doi.org/10.5281/zenodo.7897244

Lee, B. C. G. (2023). The 'collections as ML data' checklist for machine learning and cultural heritage. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24765

Lelkes-Rarugal, C. (2024). *Creating and sharing collection datasets from the UK Web Archive*. Blog post. 18th September 2024. https://blogs.bl.uk/webarchive/2024/09/creating-and-sharing-collection-datasets-from-the-uk-web-archive.html (Archived by the Internet Archive at https://web.archive.org/web/20250113042946/https://blogs.bl.uk/webarchive/2024/09/creating-and-sharing-collection-datasets-from-the-uk-web-archive.html)

Maemura, E., & Byrne, H. (2024). Datasheets for Web Archives Toolkit. *British Library*. https://doi.org/10.22020/rq8z-r112

National Library of Scotland. (2024). *National Library of Scotland AI Statement*. https://data.nls.uk/projects/ai-statement/ (Archived by the Internet Archive at https://web.archive.org/web/20250118173611/https://data.nls.uk/projects/ai-statement/)

Padilla, T., Allen, L., Frost, H., Potvin, S., Roke, E. R., & Varner, S. (2019). Always already computational. https://collectionsasdata.github.io/

Padilla, T., Scates Kettler, H., Varner, S., & Shorish, Y. (2023). *Vancouver Statement on Collections as Data. Zenodo*. https://doi.org/10.5281/zenodo.8342171

Research Data Alliance. (2024). *Collections as Data Interest Group*. https://www.rd-alliance.org/groups/collections-as-data-ig/activity/ (Archived by the Internet Archive at https://web.archive.org/web/20241217000915/https://www.rd-alliance.org/groups/collections-as-data-ig/activity/)

Ross, J., Sheail, P., Terras, M., & Schofield, C. (2024). *Infrastructure futures for digital cultural heritage*. https://doi.org/10.5281/zenodo.13710266

Scann, E., Ruiz, C., Romeo, F., and Wallace, A. (2020). Declaration on Open Access for Cultural Heritage. In *MCN Annual Conference*. https://mcn.edu/resources/declaration-on-open-access-for-cultural-heritage/ (Archived by the Internet Archive at https://web.archive.org/web/20241109091649/https://mcn.edu/resources/declaration-on-open-access-for-cultural-heritage/)

Tanner, S. (2016). Using impact as a strategic tool for developing the digital library via the balanced value impact model. *Library Leadership & Management*, 30(4). https://doi.org/10.5860/llm.v30i4.7146

Warwick, C., & Bailey-Ross, C. (2020). The impact of digital resources. In *Digital technology and the practices of humanities research* (pp. 81–104). https://doi.org/10.11647/OBP.0192.04

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). https://doi.org/10.1038/sdata.2016.18