



Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D13.2 Cultural Heritage Image Sharing Recommendations Report

Work Package	WP13 - Cultural Heritage
Lead Author (Org)	Beth Knazook (Digital Repository of Ireland)
Contributing Author(s) (Org)	<p>Joan Murphy (Digital Repository of Ireland) Keren Barner (Younes & Soraya Nazarian Library, University of Haifa) Katherine Cassidy (DRI) Steven Claeysens (KB, the National Library of the Netherlands) Claudio Cortese (4Science SpA) Eileen J. Manchester (Digital Innovation Division (LC Labs), Library of Congress) Thomas Padilla (Internet Archive) Dana Reijerkerk (Stony Brook University) Glen Robson (IIF Consortium) Antje Schmidt (Museum für Kunst und Gewerbe Hamburg) Tim Sherratt (Centre for Creative and Cultural Research, University of Canberra) Margaret Warren (Institute for Human & Machine Cognition; Metadata Authoring Systems)</p>
Due Date	31.05.2023
Date	19.05.2023
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION
DOI	https://doi.org/10.5281/zenodo.7897244

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Authors	Notes
0.9	30.04.2023	Beth Knazook, Joan Murphy	Draft for internal review
1.0	19.05.2023	Beth Knazook, Joan Murphy	Content ready

Disclaimer

WorldFAIR has received funding from the European Commission's WIDERA coordination and support programme under the Grant Agreement no. 101058393. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

Abbreviations and Acronyms

CC	Creative Commons
CDIF	Cross-Domain Interoperability Framework
DRI	Digital Repository of Ireland
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile

GLAM	Galleries, Libraries, Archives and Museums
LOD	Linked Open Data
PID	Persistent Identifier
TK	Traditional Knowledge labels
URI	Uniform Resource Identifier

Executive Summary

Although the cultural heritage sector has only recently begun to think of traditional gallery, library, archival and museum ('GLAM') collections as data, long established practices guiding the management and sharing of information resources has aligned the domain well with the FAIR principles for research data, evidenced in complementary workflows and standards that support discovery, access, reuse, and persistence.¹ As explored in the previous report by Work Package 13 for the WorldFAIR Project, D13.1 *Practices and policies supporting cultural heritage image sharing platforms*, memory institutions are in an important position to influence cross-domain data sharing practices and raise critical questions about why and how those practices are implemented.²

Deliverable 13.2 aims to build on our understanding of what it means to support FAIR in the sharing of image data derived from GLAM collections. This report looks at previous efforts by the sector towards FAIR alignment and presents 5 recommendations designed to be implemented and tested at the DRI that are also broadly applicable to the work of the GLAMs. The recommendations are ultimately a roadmap for the Digital Repository of Ireland (DRI) to follow in improving repository services, as well as a call for continued dialogue around 'what is FAIR?' within the cultural heritage research data landscape.

¹ Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3(1), p. 160018. <https://doi.org/10.1038/sdata.2016.18>

² Knazook, B., & Murphy, J. (2023) 'WorldFAIR Project (D13.1) Cultural Heritage Mapping Report: Practices and policies supporting Cultural Heritage image sharing platforms,' Zenodo. <https://doi.org/10.5281/zenodo.7659002>

Table of contents

Executive Summary	5
1. Introduction	7
1.1 Report methodology	7
1.2 Background and summary of D13.1	9
2. What is FAIR in the cultural heritage sector?	11
3. Recommendations	15
3.1 Recommendation 1 – Citation Model	15
3.2 Recommendation 2 – Transparency	18
3.3 Recommendation 3 – Data Documentation	19
3.4 Recommendation 4 – Licensing and Rights	22
3.5 Recommendation 5 – Delivery	24
4. Conclusion	26
5. Bibliography	27
Appendix A – Recommendations aligned to FAIR Principles	31
Appendix B – Mind map of thematic headings informing the recommendations in accessible list format	32
Appendix C – Image Sharing Platforms Summary Table	34

1. Introduction

Efforts made by galleries, libraries, archives and museums (GLAMs) to facilitate the discovery and use of digital image collections have already made it possible for the general public to discover much of the world's digitised cultural heritage through simple, user-friendly databases and web-based search engines, but the mechanisms by which these images are made available for research are derived from access models aimed at the general user.

In aligning these models with the FAIR principles for research data, this report seeks to provide advice to GLAM sector professionals in better delivering their collections as data for research and computational reuse. The suggestions made here are envisioned to support the collaborations and technologies already underpinning major image sharing platforms, enabling the sector to present image collections in ways that are more easily machine actionable and interoperable with other sources of data.

The following report first presents some background information on the application of FAIR principles and attempts at 'datafication' in the cultural heritage sector, before outlining 5 recommendations for increasing alignment with the FAIR principles. These recommendations are meant to be broadly applicable to collections work across the variety of collecting approaches represented by the GLAMs, reflected in the design of community image sharing platforms and aggregators, although the ultimate intention is to implement these recommendations for image collections at the Digital Repository of Ireland (DRI) within the timelines of the WorldFAIR Project.

1.1 Report methodology

Although the Cultural Heritage Case Study in Work Package 13 focuses specifically on improving FAIR supporting practices at the Digital Repository of Ireland, it was important for the project to gather feedback on the broader implications of this work from professionals in the sector with expertise in diverse areas of policy, professional practice and technology. The DRI invited participation in the Cultural Heritage Image Sharing Working Group starting in November 2022, with the first meeting taking place in January 2023. Open calls for additional membership were issued in early March to attempt to expand the global perspective on the work. In total, the group met for a period of 5 months to discuss and review the recommendations in this report.³

The proposed approach to developing and writing D13.2 was collaborative and iterative. The project team analysed feedback from working group meetings, reviewer comments on D13.1 and shared

³ Knazook, B. (2023) *Introducing the WorldFAIR Cultural Heritage Image Sharing Working Group*.
<https://worldfair-project.eu/2023/05/17/introducing-the-worldfair-cultural-heritage-image-sharing-working-group/>

notes from a webinar workshop hosted by the Research Data Alliance (RDA) in late February, and combined the results into a series of thematic headings: *Transparency, Mutability, Technology, Rights* and *Costs* (Figure 1, see also Appendix B).⁴ Members were then invited to contribute to the writing on these themes, either by attending a Zoom-based writing sprint in late March or writing asynchronously. The recommendations documents were accessible to all members for review and refinement for a period of two weeks following the Zoom events. After this point the writing was synthesised and the outputs collated by the project team into a single document, which forms the basis of this report. Figure 2 summarises the timeline of work.

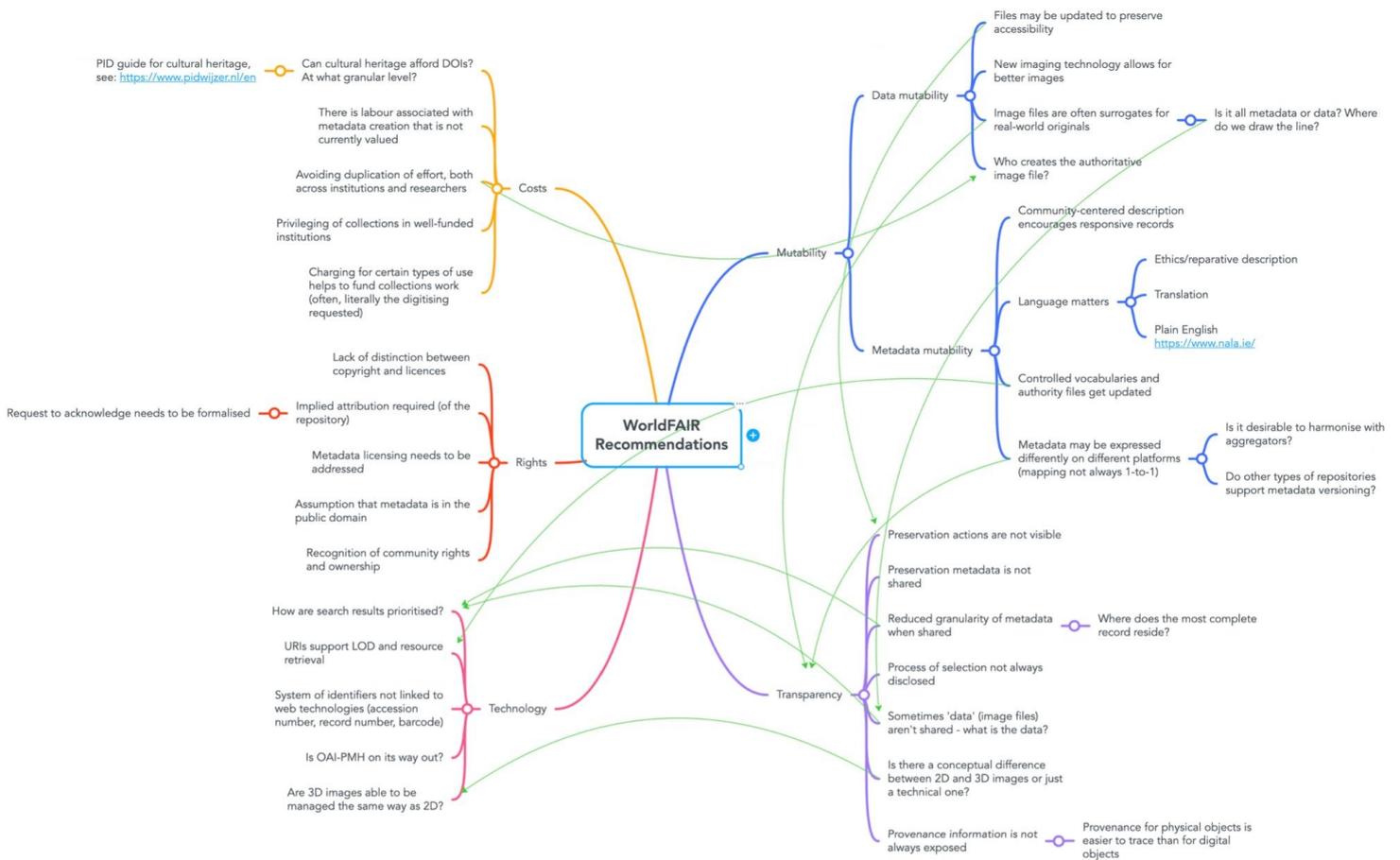


Figure 1 – Mind map of thematic headings informing the recommendations

⁴ Knazook, B. (2023) 'WorldFAIR @ RDA's 10 Year Anniversary: Image Sharing Systems and Practices in Cultural Heritage,' WorldFAIR Project. <https://worldfair-project.eu/2023/05/18/worldfair-rdas-10-year-anniversary-image-sharing-systems-and-practices-in-cultural-heritage-2/>



Figure 2 - Timeline of recommendations development

1.2 Background and summary of D13.1

D13.1 *Practices and policies supporting cultural heritage image sharing platforms* presented a landscape report of organisations and image sharing platforms in the GLAM sector, examining interoperability mechanisms through two lenses – Enabling Practices and Enabling Technologies. While we do not propose to reproduce the content of D13.1 here, a summary of our findings under these two headings may be helpful to the reader when considering the recommendations in this report.

Enabling Practices

Policies informing acquisition, selection, delivery, digitisation and preservation practices promote transparency and public accountability, and are strongly supported by professional associations and leading institutions, however these policies are not always public-facing or visibly linked to metadata records that describe GLAM collections for the end user. Persistence is supported by

stewardship obligations outlined in these policies which advocate for accessible metadata and assure some level of commitment to maintaining and keeping digital files accessible over time.

Descriptive metadata and data value standards are well-developed within the sector, though minimum compliance requirements vary across organisations and errors can arise when contributing content to image sharing platforms due to varying organisational needs and understanding. Across these platforms, Dublin Core is commonly deployed as both a descriptive and structural standard as it maps well to a variety of more complex standards, although its use also risks loss of granularity in the description and complexity in the structure of the records.⁵ When this complexity is reduced, item-level metadata that was dependent on relationships to collections-level metadata may not be as easily understood. Provenance information about the metadata itself may be lacking, especially when this metadata is copied over from institutional collection catalogues to aggregator platforms. Technical metadata pertaining to digital capture and creation is rarely exposed to the end user.

Data formats for still images are mainly homogeneous, owing to a number of established and widely adopted digitisation guidelines developed for the sector. Open formats which are well-documented, easily accessed and preserved, such as JPEG, TIFF, PNG and PDF, are almost universally agreed upon. The push for 3D imaging in cultural heritage is growing, but adoption is currently low and it is unclear if delivering 3D images to the end user will require any change in collection delivery approaches other than enabling support for 3D viewing technologies.

Copyright and licensing of images is complex, incorporating ownership, contractual restrictions and ethical considerations, making it a challenge for both the GLAMs and image sharing platforms to apply formal licences in all circumstances. Despite these challenges, there has been widespread adoption of Creative Commons licences across image sharing platforms which is helping to formalise institutional requests for attribution and address the need to make clear statements about the allowed reuse of metadata as well as the image files.⁶

Enabling Technologies

Web technologies such as HTTPS enables users to access image files on the web. HTML and XML are used to display content in a web readable format, but importantly for potential machine actionability at scale, metadata from image sharing platforms are also made available in a structured, machine readable format via APIs which allow users to customise and refine the information retrieved from collections. The OAI-PMH metadata harvesting protocol has become a key technology for exposing and sharing metadata about GLAM collections.

RDF, Linked Open Data (LOD) and Persistent Identifiers (PIDs) are at a relatively early stage of implementation across the GLAM sector, but their uptake is well supported by common practices

⁵ 'DCMI Metadata Terms,' (2020) Dublin Core. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

⁶ 'About CC licenses,' (2019) Creative Commons. <https://creativecommons.org/about/cclicenses/>

that rely on precise and specific description. GLAMs have a long history of creating identifiers to facilitate the retrieval of content, and have substantial resources of proven vocabularies and ontologies developed for analogue collections delivery from which to draw on.

IIIF (International Image Interoperability Framework) is an open standard widely supported by GLAMs, facilitating the exchange of cultural heritage images and metadata. All the information related to a digital object, including image file(s) as well as the metadata, is stored in a structured manifest implemented using JSON-LD. A IIIF implementation allows researchers to search, view and compare high-resolution images across collections and platforms, and reuse them without the need to download or duplicate materials. The potential to reduce the need for duplication, and thus the proliferation of digital content across the web, while ensuring verified integrity and access to high-quality files are key reasons why IIIF is deployed on many image sharing platforms.

Embedded metadata in image files, such as JPEG and TIFF, may include technical details about the image capture and filetype as well as descriptive metadata about the administration, preservation and content of the file, packaged in a way that is designed to travel with files that are shared. Common types of embedded metadata include the EXIF, IPTC and ICC standards. Its usage in the cultural heritage sector is inconsistent and often determined by individual content providers, although it can be tricky to retain even for those organisations that seek to capture it. Embedded metadata that is preserved in an original file can be easily lost in workflows that rely on creating derivative images for access.

D13.1 concluded with a table summarising the interoperability enabling mechanisms used by image sharing platforms which can also be found as Appendix C in this document.

2. What is FAIR in the cultural heritage sector?

The FAIR principles for research data are a deceptively simple set of guidelines envisioned to ensure the use and reuse of research data, guided by the shared belief that good data should be Findable, Accessible, Interoperable and Reusable. They have been widely endorsed by global, research data-supporting organisations (CODATA, GO FAIR, the Research Data Alliance and World Data Systems), many national governments, research funders, research-performing institutions, academic and professional associations and industry.⁷ The principles were designed to support machine actionability, described by the authors as ‘the ability of machines to automatically find and use data,’ and to encourage data reuse across academic disciplinary boundaries, and to do so, they mainly advocate for structured and predictable information about how data are generated, curated

⁷ See, in particular, the industry example presented in Romacker, M. (2022) ‘From application-centric to data-centric using FAIR architectures and resources,’ EOSC Symposium 2022 [presentation].

<https://symposium22.eoscfuture.eu/symposium/keynotes/>

and shared.⁸ They have arguably begun to produce a language of FAIR data exchange, formed through the proliferation of familiar and consistent examples of *what a FAIR compliant dataset should look like* (e.g. a package stored in a data repository containing at minimum the data, documentation, attribution, a pathway to retrieval and a licence for reuse), encouraged by data repository design and FAIR-supporting guidance.

The proliferation of FAIR assessment tools which help to guide researchers in the preparation of datasets (e.g. ANDS-Nectar-RDS, F-UJI) have also helped to reduce the ambiguity of ‘what is FAIR?’ through straightforward questions detailing how the 15 guiding principles behind Findability, Accessibility, Interoperability and Reusability may be achieved, practically and technically. The *Turning FAIR into Reality* report asserts that services which support FAIR should ideally be FAIR themselves, and so similarly offers repositories a set of general steps to take in operationalising FAIRness:

- create a policy for deploying PIDs,
- insist on minimum metadata, ideally with the use of semantic terms,
- provide clear usage licences,
- and use well-established communications protocols like HTTP and HTTPS.⁹

At a glance, many GLAM image collections would appear to be largely compliant with FAIR: URIs, if not PIDs, are used for retrieval of both aggregate collections and individual items; metadata standards are widely used for both metadata structure and description; rights information and usage advice is generally offered; and images and metadata are typically delivered via HTTP or HTTPS. Large-scale, computational research is facilitated by the increased sharing of metadata records in machine readable formats (e.g. XML, JSON) and by the development of open APIs, which allow users to customise the parameters of collection queries. Yet cultural heritage data has proven challenging for scholars to conceptualise as data inputs and outputs within this ecosystem, particularly for those working within the arts and humanities,¹⁰ in part because although images may be accessed individually online, they are largely collected, described and managed in aggregate (i.e. items are generally acquired and interpreted as part of collections), but perhaps also because the idealistically generalisable model for FAIRness encourages thinking about humanities data as scientific data. As Erzsébet Tóth-Czifra explains,

⁸ Wilkinson, M.D. *et al.* (2016) ‘The FAIR Guiding Principles for scientific data management and stewardship’, *Scientific Data*, 3(1), p. 160018. <https://doi.org/10.1038/sdata.2016.18>

⁹ European Commission, Directorate-General for Research and Innovation (2018) *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office, p.52. <https://data.europa.eu/doi/10.2777/1524>

¹⁰ See, for example, the section ‘What does our ‘data’ look like?’ in Cridford, T. and Walker, C. (2020) *Painting the road to Research Data Management at the RCA* at RDMF20: RDM and Data Sharing/Openness in The Arts - Virtual Forum, London, UK. <https://researchonline.rca.ac.uk/id/eprint/4411>

contrary to their general scope and deliberately domain-independent nature, [the FAIR principles] have been implicitly designed according to underlying assumptions about how knowledge creation operates and communicates... first, scholarly data or metadata is digital by nature, second, scholarly data is always created and therefore owned by researchers, and third, there is a wide community-level agreement on what can be considered as scholarly data.¹¹

These issues manifest as tangible barriers to aligning cultural heritage resource sharing fully with FAIR. Many GLAM organisations act as stewards for collections that they do not own and rely on agreements that grant them permission to share, but not necessarily to extend that permission to others. There may be additional difficulties in assigning licences to materials which require consideration of other ownership rights, such as those implied by Indigenous data sovereignty. Compounding these concerns are the relative scarcity of digitised resources compared to the volume of cultural heritage materials in collections and a corresponding lack of capacity to provide robust technical and contextual description. Proposed FAIR supporting actions such as ‘provide more provenance metadata’ and ‘use CC licences’ might not be reasonably achieved by all. There are additional ethical obstacles complicating the process of alignment, due to an absence of recognition from FAIR assessment guides of the importance of multilingualism, community representation, timeliness and change in descriptive ethics, as well as the limitations to technical and financial investment in producing digital data.

Still, a few efforts have been made in recent years to articulate how cultural heritage collections may be reimagined in the data economy. In 2018, Koster and Woutersen-Windhower produced a set of FAIR recommendations for (G)LAM institutions in which they asserted ‘it is not enough to make collections available through web based end user interfaces and provide download options for individual objects and metadata records.’¹² Despite that their recommendations directly responded to the FAIR principles as articulated by Wilkinson *et al.*, they also gave an early indication that FAIR was never likely to work for the cultural heritage sector as is: ‘The two main limitations are the lack of explicit attention for long term preservation of digital objects, besides their metadata, and the excessive interwovenness of “data” (or objects) and “metadata”.’ The former concern has since been addressed by the TRUST principles for digital repositories, which advocate for reliability, transparency and accountability in the repositories that store and provide access to data.¹³ The problem of distinguishing the nuances of the guidance for metadata and data, however, remains an issue not just for the cultural heritage sector, but for many of the disciplinary case studies within the WorldFAIR Project.

¹¹ Tóth-Czifra, E. (2019) ‘The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem,’ *HAL*. <https://shs.hal.science/halshs-02115505>

¹² Koster, L. & Woutersen-Windhower, S. (2018) ‘FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections,’ *Code4Lib Journal*, 40. <https://hdl.handle.net/11245.1/4ee73567-5ddd-4bbe-9188-173ad5b0f18b>

¹³ Lin, D., Crabtree, J., Dillo, I. *et al.* (2020) ‘The TRUST Principles for digital repositories,’ *Sci Data* 7, 144. <https://doi.org/10.1038/s41597-020-0486-7>

A special issue on the topic of FAIR for the *International Journal of Digital Libraries* suggested that ‘Reuse’ needed to be expanded into ‘Reuse, Relevance and Reliability’ to properly document the origins of data alongside the institutional and intellectual frameworks that produce and manage them.¹⁴ Initiatives such as *Always Already Computational: Collections as Data* and its successor, *Collections as Data: Part to Whole*, have encouraged the GLAMs to begin to work towards consensus around building support for advanced computational research methods in the sector.¹⁵ The ‘Santa Barbara Statement on Collections as Data’ concludes with a recommendation to embrace a certain amount of messiness, advising ‘that the development of collections as data is an ongoing process and does not necessarily conclude with a final version.’¹⁶

An overarching theme that has emerged from the Cultural Heritage Case Study is the perception that the GLAMs, and by extension the FAIR principles applied to GLAMs, should support and encourage a responsive data practice that recognises cultural heritage images and their metadata will change over time, requiring additions to the interpretive frameworks and access pathways provided. The problems which Koster and Woutersen-Windhouver identified persist today in the approach to cultural heritage images as surrogate records for reference and display, rather than data to be questioned, rearranged and reused. Interoperability is here envisioned as opening the lines for communication between humans engaging with records of cultural memory, supported by machines at any scale, which is ultimately better realised as a dialogue rather than a service. The following recommendations strongly encourage the foregrounding of human perspectives and interpretive contexts which inform the process of sharing images as data, made practical through approaches that leverage existing practices in the GLAMs.

¹⁴ Hermon, S. & Niccolucci, F. (2021) ‘FAIR Data and Cultural Heritage Special Issue Editorial Note,’ *International Journal on Digital Libraries*, 22, 251-255. <https://doi.org/10.1007/s00799-021-00309-8>

¹⁵ Padilla, T., Scates Kettler, H., Shorish, Y., Varner, S. (2023) *Collections as Data – Part to Whole*. <https://collectionsasdata.github.io/part2whole/>

¹⁶ Padilla, T. *et al.* (2018) ‘Santa Barbara Statement on Collections as Data – Always Already Computational: Collections as Data,’ Zenodo. <https://doi.org/10.5281/zenodo.3066209>

3. Recommendations

There are 5 recommendations presented here. For each, there is a ‘Rationale’ provided which highlights the reasons why the recommendation is needed. This is followed by a ‘FAIR alignment’ statement which suggests how FAIR assessments may acknowledge this as a FAIR-aligned practice. Below these sections is a ‘Background discussion’ that elaborates on how we came to the recommendation. The authors acknowledge that there is overlap between many of these recommendations, as evidenced by returning themes in the supporting discussions.

3.1 Recommendation 1 – Citation Model

1.	A formal citation model for cultural heritage images should be adopted which recognises digital surrogates as research objects and includes references to revisions of either image data or metadata
Rationale	Cultural heritage organisations should adopt a citation model that draws attention to the fact that metadata and files maintained by GLAM organisations will change over time. In order to ensure a transparent and trustworthy record, the date of creation of a digital file, date of publication and the date(s) of any revisions to metadata should be clearly stated so that researchers can accurately cite the image, including its contextual framing and relationship to other image iterations, as it existed at the point in time they accessed it. It may also be advised to include any identifying numbers issued by the institution that have been historically used to access the material. The GLAMs are further encouraged to expose records that document the rationale for changes to metadata or data, and expose reparative description work that may have been carried out in the past.
FAIR alignment	While no specific citation model is proposed by the FAIR principles, popular choices for disciplinary and generalist repositories follow citation models devised for publications. These focus on a limited number of metadata points: authorship, date, publisher and location. It is common for research data repositories to provide a version number and date as well, with each version sometimes also assigned a unique PID that links to a preserved record of current and prior versions. This is not necessarily possible for all cultural heritage images, as it presents a potentially substantial cost to collecting institutions to maintain multiple versions of data or metadata, while also risking perpetuating potentially problematic narratives by retaining metadata that should be recontextualised. The critical piece for alignment seems to be that the citation refers to a uniquely identifiable object and alerts the researcher to the presence of changes made

	over time, not necessarily that all versions of the data remain available in perpetuity.
--	--

Background discussion:

The Working Group agreed broadly that the GLAM sector has a conceptual problem to overcome in the assumption that digital representations of images are mere surrogates for original objects. The digital files made available on image sharing platforms are unarguably primary research objects in and of themselves, and information about those objects is important to communicate. But generally speaking the choice of metadata fields, and the way they are used in combination in citation styles favoured by the GLAMs, typically foreground descriptive details that support an interpretation of the original object and are less useful to the retrieval of specific digital representations. In reality, many citations are simply descriptive captions. Take this example from the Digital Public Library of America (Figure 3):



Cite this item

Nourse, Henry King, 1880-1968. [Cat]. [between 1900 and 1919]. Retrieved from the Digital Public Library of America, https://csl.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma990014663490205115&context=L&vid=01CSL_INST:CSL. (Accessed May 16, 2023.)

Figure 3 - Image of cat on DPLA site, with citation (Source: <https://dp.la/item/d9b88909f665eaaaca7054ffa1fe2d79?q=cat>)

The citation provided is essentially a combination of descriptive metadata fields that have been stripped of their labels and rearranged into a format compliant with the *Chicago Manual of Style*.¹⁷ The title describes the main subject of the resource (in this case, a cat), there is an author credit, and the time period seems to pretty clearly indicate the date of creation of the original object rather than the date of digitisation or publication in the repository. The citation concludes with an acknowledgement of the DPLA repository as the source and yet the URI for retrieval of the digital access image points to the catalogue of the originating collection (in this case, the California State Library). There are technically 3 different objects being described by one citation: the original analogue photograph, the metadata record in the DPLA and the digital object held by the California State Library. A citation style is needed that corrects this confusion of sources and provides information to support the retrieval of the digital asset, the institutional framework that produced it and which acknowledges the likelihood of multiple representations and the relationships between them. Adding date fields related to the digital image file addresses the time dimension, while additional metadata that situates the research practice of the institution and collections context could also be explored.

When framing collections as data, the GLAM sector should seek to foster practices associated with first-class research outputs such as the use of unambiguous citations. The *Joint Declaration of Data Citation Principles* state that ‘Data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.’¹⁸ The development of a standardised citation model for cultural heritage is currently being explored by researchers supported by an EOSC Future grant, with a proposed delivery date of September 2023. Viewing citation as a pillar of Open Science, the project aims to address ‘the basic citation requirements that can identify the citationability of the digital representation of the cultural artefact and ensure derivation from the adopted data model to certify its trustworthiness in the terms proper to the FAIR guiding principles of Traceability, Accessibility, Interoperability, Reusability.’¹⁹ The DRI project team looks forward to refining our citation style with input from this project.

DPLA was alone amongst the image sharing platforms to provide a citation of any kind, and it should be noted that the issue of citation was not examined in D13.1, although it has been added retrospectively (information reflected in Appendix C).

¹⁷ These metadata fields can be combined in different ways on the DPLA website to conform with different citation standards. This output is formed to match the *Chicago Manual of Style*, but *APA* and *MLA* citations were also output options. None of the metadata provided aligns exactly with the fields requested by these standards however, and DPLA acknowledges that ‘These citations are programmatically generated and may be incomplete.’

¹⁸ Data Citation Synthesis Group (2014) *Joint Declaration of Data Citation Principles*. Force11.

<https://doi.org/10.25490/A97F-EGYK>

¹⁹ Biblioteca Digitale Padova, ‘Library System Projects’, Biblioteca Digitale Padova – Università di Padova,

<https://bibliotecadigitale.cab.unipd.it/en/digital-library/library-system-projects>

3.2 Recommendation 2 – Transparency

2.	Information about the creation, management and preservation of files should be visible and understandable by both humans and machines
Rationale	This recommendation builds on the prior one in asking for documentation describing changes to files and the relationships between files, objects and collections, recognising that some effort to capture large-scale, historical change within GLAM collections is critical to their use as research assets. Digital objects may be replaced over time to ensure ongoing compatibility with changing platforms and technologies. Whether preservation actions are carried out purely to maintain the file format or to improve the user experience with better digital capture technology (the re-digitisation of objects), any action that changes the image file presented to the end user should be clearly communicated as part of the available metadata. This may be realised through embedded technical metadata, metadata provided at the item level, or collection-level metadata that broadly documents digitisation and preservation efforts.
FAIR alignment	In the sciences, instrument data is augmented by information about the instrument, which is no different than providing the user with digital capture information in digitised or born-digital collections material. A particular lens in a particular camera may impart defects or represent the source material differently from another camera, which could influence both human and mechanical interpretation. As data provenance is also key for interoperability and reuse, the provenance of file types generated by the institution in the course of preservation work should be prioritised to support ongoing computational use. This may come in the form of technical metadata that is readily available and embedded in image files, or it may be satisfied by collection-level metadata that describes the digitisation and/or digital preservation process. The level of preservation provided by the institution will influence the amount and type of metadata made available. ²⁰

Background discussion:

Digitised objects are often envisioned as windows onto the ‘real life’ objects they depict, which has meant the sector has not always prioritised the sharing of technical metadata. Yet the potential computational reuse of digital files is, in the end, inextricably linked to the digital data provided by

²⁰ See, for example, National Digital Stewardship Alliance (2019) ‘Levels of Digital Preservation, Version 2.0’ <https://ndsa.org/publications/levels-of-digital-preservation/>

the camera, the filetype and the management actions that ensure ongoing access. Only by knowing the technical metadata is it possible, for example, to understand how well a digital image may conform to the original cultural asset.

The Dublin Core metadata standard, identified in D13.1 as the predominant standard in use by the GLAM sector, was originally developed in response to a need to productively manage digital library collections with appropriate metadata that described their technical format (JPEG, TIFF, PNG, etc.), type (still image, video, and so on) size (GBs, pixels) and source (via unique identifiers and links), as well as their conceptual form (subject, title and description). In reality, the flexibility of the standard has blurred the lines between the analogue and the digital, with the digital information often ignored.

There was some concern expressed within the Working Group around the limitations of sharing and exposing the technical metadata of digital cultural resources because of potential financial costs in obtaining this metadata from third-party digitisers or costs associated with additional data storage locally, which may mean that some institutions opt not to preserve this fuller metadata at all. Problems were also noted with workflows that inadvertently strip access files of the technical information embedded in the originals. In situations where Dublin Core or similar metadata standards have been deployed without capturing this information, it may not be possible to go back and enhance individual image records. One remedy could be to expose the full metadata capture of preservation actions recorded by the Preservation Metadata Implementation Strategies (PREMIS) standard where this detail is available, but there is a risk of overwhelming the user with information events that document minor management activities like fixity checks, and not all institutions have retained the same level of information around preservation actions. Managing different types of metadata in different ways negates the risk of overwhelming the user.

3.3 Recommendation 3 – Data Documentation

3.	The process of selection, scope and completeness of the data should always be documented
Rationale	GLAM institutions may have rigorous processes for selection, but the rationale for specific acquisition and digitisation decisions is rarely communicated to the end user on a collection-by-collection or even an individual record basis. Exceptions exist for collections that are still being processed or when accruals are expected, as this is usually documented in the public-facing record metadata. We recommend that additional information be shared wherever possible, about the selection process and expected additions through future digitisation or acquisition plans, particularly in cases where there have been omissions or other

	<p>considerations defining the boundaries of collections currently available online. This may be included in collection-level metadata, and could be indicated at the item level where appropriate.</p>
<p>FAIR alignment</p>	<p>In research projects, the scope of the research question is often included as part of the documentation so that users understand not only what is available within the dataset, but the selection process that produced that data, whether curated from a larger dataset or compiled from multiple sources. GLAM organisations rarely ‘package’ collections in ways that completely document the history of the object in the organisation, instead providing links that indicate relationships to other collections materials, policies guiding decision-making, or statements about acquisition and selection at the aggregate level. Here the production of data documentation designed to provide a parallel to a research dataset’s README file is recommended, but any documents provided which share institutional perspectives should be acknowledged as fulfilling this need.</p>

Background discussion:

There is growing recognition among the GLAMs that they influence research agendas by collecting and making certain materials for research available, and by providing metadata framed by historical collecting practices and biases. It is widely acknowledged that there is an obligation to communicate to their publics how collections are acquired, and why some collections objects may be made available and not others, but the communication strategies often rely on high-level documents or records that might only be made available internally. Many users assume that GLAMs are ‘digitising everything’ and what is represented online is an unbiased and balanced representation of the whole scope of historical data available. There is an increased acknowledgement amongst GLAM professionals that the acquisition, processing and digitisation of collections depends on collections mandates, outreach programmes, availability of materials, copyright, embargoes or contractual agreements, funding, research requests and considerations for equitable representation.

Cultural heritage is further unique among the disciplines in that the metadata created to describe research objects is typically not provided by a researcher, author or creator pursuing a research inquiry, nor a machine fine-tuned to measure, observe or produce calculations and analyses. There may be many people involved with different expertise in the processing, description or digitisation of collections acting independently at various stages, both during the initial acquisition and cataloguing phase and at potential points of record enhancement and re-description over the life of the object in the collection. As it is likely that AI generated metadata will become increasingly

common within the sector, this too will require consideration.²¹ Given the range of people involved in the processing of cultural heritage data prior to publication on a public-facing website many of decisions are not transparent to the end user, with connections not always being made to existing, publicly available collection development policies. Extending the scope of policy to frame decisions at the data level would be helpful in this regard.

There is also often no associated publication explaining and defending the knowledge frameworks guiding the presentation of metadata, though exhibitions and other narrative tools may be made available. A critical issue that the field is currently grappling with is that both data and metadata sometimes reproduce and disseminate problematic historical narratives necessarily entangled within the institution's historical perspective. The stories about how collections came to be in memory institutions rarely form a part of the official record, but can be learned through conversations with archivists, librarians, curators and collections managers. If any GLAM collection images are to be truly machine interoperable, the narratives that surround the images must be better documented so that the researcher seeking to reuse the data knows what methodology or critical interpretive framework influenced its presentation.

Practical examples of how this may be addressed could be by leveraging and/or extending existing search indexes to provide some of this contextual information, e.g. by letting users filter/facet on a 'date modified' field, or obtain data about the number of images on a topic by year of creation, etc. Another approach could be to expose and/or capture statistical information about collections as a whole to help researchers to understand how a particular results set relates to the whole.

Datasheets for datasets have emerged as an important tool for increasing transparency and accountability within the machine learning community, encouraging data creators to reflect 'on the process of creating, distributing, and maintaining a dataset, including any underlying assumptions, potential risks or harms, and implications of use' in order to ensure data consumers 'have the information they need to make informed decisions about using a dataset.'²² The datasheet model proposed by Gebru *et al.* (2021) comprises a set of questions which can be adapted to a particular domain of research, and this work is already being piloted by some GLAMs. The Library of Congress 'LC Labs' tested a document they called a 'cover sheet' for datasets as part of the Computing Cultural Heritage in the Cloud Data Jam.²³ Seven expert users were invited to analyse three data packages, and they all indicated they were very interested in the acquisition, digitisation, and

²¹ Heus, P. (2023) 'AI has a metadata problem...', *Medium*, 24 April.

<https://plgah.medium.com/ai-has-a-metadata-problem-78b30ca1936b>

²² Gebru, T. *et al.* (2021) 'Datasheets for Datasets,' *Communications of the ACM*, 64, 12, (86-92).

<https://doi.org/10.48550/arXiv.1803.09010>

²³ Manchester, E.J. (2022b) 'Now Playing: the CCHC Data Jam!' *Library of Congress Blogs*.

<https://blogs.loc.gov/thesignal/2022/12/now-playing-the-cchc-data-jam/>

descriptive history of the items.²⁴ Similar work is being undertaken to develop datasheets at Europeana, indicating a need to reconsider the ways in which GLAMs contextualise image records.²⁵

3.4 Recommendation 4 – Licensing and Rights

4.	Rights, licences and labels should be applied consistently and in a standardised way
Rationale	The use of Rights Statements ²⁶ to formalise and make clear the legal status of both original objects and digital representations is recommended. We further recommend the application of the Creative Commons CC0 ‘no rights reserved’ licence for digital image metadata as the default standard, recognising that there may be considerations in certain circumstances requiring the use of a more restrictive licence. ²⁷ Applying formal licences for metadata recognises that the institution is often the author of the records, and supports the interoperability and reuse of information about images regardless of the complexity of relationships that may inform the licensing of the image file itself. Where appropriate, the array of creators, contributors and communities may be recognised by the addition of labels.
FAIR alignment	Explicit rights and licensing statements are a requirement for all aspects of the FAIR data workflow, from discovery to reuse, and this information should be available in both human readable and machine actionable forms. The application of standardised rights statements for digital files is probably far more feasible within the sector than machine actionable licences however, and should be acknowledged as good practice. Despite the complexity of this work, recognising other forms of attribution, ownership and care should also be prioritised.

Background discussion:

As the GLAM sector increasingly moves to view collections as data, there is a need to make explicit the copyright and licences inherent in both the metadata and data itself. Often cultural heritage institutions do not apply licences to their collections at all, or they make requests that have unclear

²⁴ Manchester, E.J. (2022a) ‘Announcing LC Labs Data Sandbox and 3 New Data Packages,’ *Library of Congress Blogs*. <https://blogs.loc.gov/thesignal/2022/12/announcing-lc-labs-data-sandbox-and-3-new-data-packages/>

²⁵ Europeana (2023) EuropeanaTech Workplan 2023. <https://pro.europeana.eu/post/europeanatech-work-plan-2023>

²⁶ ‘Rights Statements’ (n.d.) RightsStatements.org. <https://rightsstatements.org/en/>

²⁷ ‘CC0 – Creative Commons,’ Creative Commons. <https://creativecommons.org/share-your-work/public-domain/cc0/>

licensing implications.²⁸ They may charge fees for the re-use of collections and negotiate exclusive agreements with industry partners for digitisation.²⁹

An additional challenge for the sector is that objects may themselves be in the public domain, while also being part of institutional collections, this is an issue that may be addressed by the OpenGLAM Principles.³⁰

Cultural heritage data and metadata have additional considerations when it comes to the application of usage licences which may not be well-addressed by one-size-fits-all frameworks widely used in other fields of research (for example, Creative Commons licences). There are often problems of distinction between a cultural heritage object, the digital surrogate of a cultural heritage object and access versions of the same, which can lead to confusion in applying licences. Correctly identifying the creator or copyright owner is especially confusing when the cultural heritage object might contain layers of potential attributions, particularly when different images document or depict the same original cultural object. If it is unclear who may be able to issue a licence, the copyright status of the work is probably also unclear.

When we transform collections into data, ethical questions are raised that are not adequately covered by FAIR. The CARE principles emerged as a complementary guide to FAIR in order to acknowledge the rights of Indigenous communities in the representation of data deeply tied to their history, land and persons, and which are not otherwise given authority in the knowledge systems inherent in FAIR.³¹ Community ownership may not be adequately addressed by licences, which makes it problematic to recommend further compliance in this way. Using community-sourced tags or labels embedded into the metadata record is probably the best alternative at present. For institutions that wish to make use of community-owned and generated context labels, such as the Traditional Knowledge (TK) labels, Notices can be generated via the Local Contexts Hub for a variety of platforms (including repositories, datasets and more) 'to identify Indigenous collections and data and recognize Indigenous rights and interests.'³²

Changing this landscape is a much bigger task than can be accomplished within the scope of the WorldFAIR Project, but it does seem possible to make at least one suggestion to improve the way the sector expresses this information. It seems reasonable to ask that metadata be released under a CC0 licence or 'no rights reserved' statement where possible, acknowledging that metadata is often generated at multiple tiers, including descriptive, technical and preservation, and is created by

²⁸ Knazook, B., & Murphy, J. (2023) 'WorldFAIR Project (D13.1) Cultural Heritage Mapping Report: Practices and policies supporting Cultural Heritage image sharing platforms,' Zenodo. <https://doi.org/10.5281/zenodo.7659002>

²⁹ Tasovac, T., Chambers, S. & Tóth-Czifra, E. (8 October 2020) 'Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper,' HAL. <https://hal.archives-ouvertes.fr/hal-02961317>

³⁰ OpenGLAM Working Group (n.d.) 'OpenGLAM Principles,' OpenGLAM. <https://openglam.org/principles/>

³¹ Global Indigenous Data Alliance (n.d.) *CARE Principles*, Global Indigenous Data Alliance. <https://www.gida-global.org/care>

³² 'TK Labels' Local Contexts (n.d.) <https://localcontexts.org/notices/aboutnotices/>

numerous actors across the data’s lifecycle. Where metadata may be commercially sensitive or of a proprietary nature, it is even more important that it be clearly identified as such with a formal licence statement.

3.5 Recommendation 5 – Delivery

5.	Improved support for identifier schemes, APIs and machine-readable contextual data
Rationale	If collections are truly to function as data, elements within a metadata record should be independently interpretable and connected to related topics, content and terminology, facilitated through the use of linked data vocabularies, and using standard protocols and frameworks.
FAIR alignment	FAIR recognises the value of PIDs in aiding discovery, accessibility and interoperability of data and metadata, but stable links of any kind to the semantic web of data has a potentially significant implication for reusability as well. LOD offers connections to concepts which aid the understanding of ideas in multiple languages, and help to tell the stories of creators, owners and communities that inform those connections. Standardised frameworks enable interoperability by presenting structured machine-actionable data.

Background discussion:

The (re)usability of data is dependent on users being able to link to datasets and ensure the links keep working. Guaranteeing that URIs will remain persistent is one of the biggest challenges in long-term digital archiving, and yet URIs are widely used in Linked Open Data to support the retrieval of standardised descriptions. Can URIs rely on the host remaining stable? Although a PID service would be an ideal end-game, PIDs for cultural heritage images may prove too costly for many institutions to utilise. Finding an approach that ensures a persistent combination of information (i.e., institution name + a unique local identifier, or URI and related stable metadata links from controlled vocabulary resources like the Virtual International Authority File (VIAF)) may prove to be a reasonable alternative. There is also some promising work ongoing with newer technologies being developed in different file formats which use hash identifiers stored via proof-of-storage type blockchain techniques (e.g. IPFS), perhaps making the technical question of how to provide links less important than what should be reliably described by them.

It may be time to consider that preserving a link to a particular metadata record at a particular point in time is not the only way in which to make things findable. PIDs, citations and licences for datasets reliably return files and their associated information, but they also presume that the stories data tells are able to be told without an abundance of voices. The use of LOD has the potential to both add complexity through the use of agreed-upon vocabulary terms and authority metadata, and can also be used to build relationships which put the metadata in dialogue with emerging ideas and practices. LOD resources are often provided by established collecting institutions, but they may also be published by researchers, communities or smaller organisations, which should be encouraged. Aggregators like the Social Networks and Archival Contexts (SNAC) cooperative have shown what diverse open links to archival authority information can do to expand the scope of knowledge on creators and collections.³³ There are other tools emerging as well which use LOD to resolve entities through public datasets such as DBpedia, Art and Architecture Thesaurus, Wikidata and the CIDOC-CRM, storing data in RDF, with the metadata output guided by the FAIR principles.³⁴

LOD can also address an issue with choice of language in a metadata record. Cataloguers trained to use controlled vocabularies will often select the narrowest term available from within a given vocabulary, to add precision to the user browsing experience, and the relationship of those terms to hierarchical or cross-linked systems of knowledge can help to more clearly describe content for non-specialists and link to the term in other languages for non-native speakers. Members of the Working Group observed that English translations were often desirable for contribution of metadata to aggregators and collaboration with non-English organisations, which can be aided by multilingual LOD resources.

The addition of technical metadata fields suggested in Recommendation 2, such as ‘date modified’, would help researchers understand how an item has changed over time and these fields can also be used as filters/facets in a search /API request, making it possible to generate a snapshot of a collection at a particular point in time. Similarly, depending on available fields, snapshots could be generated that show patterns in digitisation, copyright or the application of specific licences. These snapshots, ideally saved at regular intervals into a version control system, would provide a big-picture of the collections that researchers could use to contextualise the results of their own search queries. Some effort to capture large-scale, historical change within GLAM collections is critical to their use as research assets.

Protocols and frameworks, when widely adopted, offer a means to provide compliance with many of the elements of the FAIR principles, and while this report aims to be technology agnostic, mention should be made again of the International Image Interoperability Framework (IIIF) identified in our previous report as a key technology to aid the discovery of images. An open source community standard widely supported by leading institutions and bodies in the GLAM sector, IIIF addresses a number of the challenges identified within GLAMs and offers a solid framework for the

³³ Social Networks and Archival Context (n.d.) ‘About SNAC.’ <https://portal.snaccooperative.org/about>

³⁴ Metadata Authoring Systems (n.d.) ‘ImageSnippets.’ <https://imagesnippets.com/index2.html>

identification, description, delivery and retrieval of image metadata. Many functionalities align with FAIR principles such as the requirement for unique identifiers, open access to metadata via the IIF Manifest, and an agreed standard API to allow multiple tools and collections to access and use the data.³⁵

4. Conclusion

As shown in the D13.1 report, the Digital Repository of Ireland's policies and practices are largely in line with common practice observed in the cultural heritage sector, and the current repository design reflects similar technical and conceptual choices seen in a variety of popular image sharing platforms. It supports both researchers who wish to deposit datasets and collections managers looking to preserve collections material that may form the basis of a research inquiry. For the DRI, these recommendations must be able to be implemented repository-side, in terms of technology and policy development, while being mindful of the needs of the membership the repository serves. It is considered important to avoid creating significant new workflows and additional workload across the members. The goal of this work is to create approachable and sustainable methods to support cross-domain use of GLAM image collections as data, and to advocate for ongoing collaboration within the field.

The next steps for the DRI will be to meet with a range of internal stakeholders to agree an approach to implement the recommendations outlined in this report. The approach aims to be collaborative, seeking to reach agreement by consensus while remaining at all times cognisant of the daily demands on both the repository's people and its technical capabilities. Implementation planning will begin in June, and given the DRI's status as CoreTrustSeal certified repository, it is anticipated that there will be an initial focus on policy and process requirements (Figure 4).

Both the processes and outputs of the implementation phase will be documented in detail and published in the final deliverable D13.3, providing a reproducible roadmap for other organisations who wish to implement the recommendations identified in this report and embed FAIR principles into their work practices.

³⁵ Padfield, J., Delaney, O. & Alberti, A.A. (2022) 'D5.16 Report on making heritage science data FAIR (Open data in Heritage Science and Archaeology)'. <https://zenodo.org/record/6779595>



Figure 4 - High level timeline for next steps

5. Bibliography

‘About CC licenses,’ (2019) Creative Commons. <https://creativecommons.org/about/ccllicenses/>

Barner, K. (2022) ‘The Creation of Primary Sources Digital Collections in an Academic Library’, *Library Philosophy and Practice (e-journal)* [Preprint]. <https://digitalcommons.unl.edu/libphilprac/7535>

Bharati, A. *et al.* (2019) ‘Beyond Pixels: Image Provenance Analysis Leveraging Metadata’, *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1692–1702. <https://doi.org/10.1109/WACV.2019.00185>

Biblioteca Digitale Padova, ‘Library System Projects’, Biblioteca Digitale Padova – Università di Padova, <https://bibliotecadigitale.cab.unipd.it/en/digital-library/library-system-projects>

Bikakis, A. *et al.* (2021) ‘Editorial: Special issue on Semantic Web for Cultural Heritage’, *Semantic Web*. Edited by P. Hitzler *et al.*, 12(2), pp. 163–167. <https://doi.org/10.3233/SW-210425>

Buttegieg, P.L. (2023) ‘WorldFAIR Project (D11.1) An assessment of the Ocean Data priority areas for development and implementation roadmap’, Zenodo. <https://doi.org/10.5281/zenodo.7682399>

Candela, G. *et al.* (25 October 2022). *Towards implementing Collections as Data in GLAM institutions*. Zenodo. <https://doi.org/10.5281/zenodo.7789480>

Carboni, N. and de Luca, L. (2017) 'Towards a Semantic Documentation of Heritage Objects through Visual and Iconographical Representations', *International Information & Library Review*, 49(3), pp. 207–217. <https://doi.org/10.1080/10572317.2017.1353374>

'CC0 – Creative Commons,' Creative Commons. <https://creativecommons.org/share-your-work/public-domain/cc0/>

Chambers, S. *et al.* (2023) 'Position Statements -> Collections as Data: State of the field and future directions.' <https://doi.org/10.5281/zenodo.7897735>

Cridford, T. & Walker, C. (2020) *Painting the road to Research Data Management at the RCA*, RDMF20: RDM and Data Sharing/Openness in The Arts - Virtual Forum, London, UK. <https://researchonline.rca.ac.uk/id/eprint/4411>

'Cultural Heritage Data Reuse Charter: the Mission Statement' (2018) *Heritage Data Reuse Charter*. <https://datacharter.hypotheses.org/77>

Data Citation Synthesis Group (2014) 'Joint Declaration of Data Citation Principles,' Force11. <https://doi.org/10.25490/A97F-EGYK>

'DCMI Metadata Terms,' (2020) Dublin Core. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

Dutch Digital Heritage Network, 'Persistent Identifier Guide,' <https://www.pidwijzer.nl/en>

European Commission (2021) *Commission proposes a common European data space for cultural heritage | Shaping Europe's digital future*. <https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage>

European Commission, Directorate-General for Research and Innovation (2018) *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office. <https://data.europa.eu/doi/10.2777/1524>

Europeana (2023) *Research Community Workplan*. <https://pro.europeana.eu/post/europeana-research-community-work-plan-2023>

Gebu, T. *et al.* (2021) 'Datasheets for Datasets,' *Communications of the ACM*, 64, 12, (86-92). <https://doi.org/10.48550/arXiv.1803.09010>

Global Indigenous Data Alliance (n.d.) *CARE Principles*, *Global Indigenous Data Alliance*. <https://www.gida-global.org/care>

- Grant, R. (2021) 'Recordkeepers Managing Research Data: a Survey of Irish Organisations,' *International Journal of Digital Curation*, 5, 1, (2-9). <https://doi.org/10.2218/ijdc.v15i1.693>
- Gregory, A. & Hodson, S. (2022) 'WorldFAIR Project (D2.1) FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' <https://doi.org/10.5281/ZENODO.7378109>
- Hermon, S. & Niccolucci, F. (2021) 'FAIR Data and Cultural Heritage Special Issue Editorial Note,' *International Journal on Digital Libraries*, 22, 251-255. <https://doi.org/10.1007/s00799-021-00309-8>
- Heus, P. (2023) 'AI has a metadata problem....,' *Medium*, 24 April. <https://plgah.medium.com/ai-has-a-metadata-problem-78b30ca1936b>
- KBR (2022) 'Linked Data and International Standards for Cultural Heritage,' KBR Conference. <https://www.kbr.be/en/agenda/linked-data-and-international-standards-for-cultural-heritage/>
- Knazook, B. (2023) 'WorldFAIR @ RDA's 10 Year Anniversary: Image Sharing Systems and Practices in Cultural Heritage,' WorldFAIR Project. <https://worldfair-project.eu/2023/05/18/worldfair-rdas-10-year-anniversary-image-sharing-systems-and-practices-in-cultural-heritage-2/>
- Knazook, B., & Murphy, J. (2023) 'WorldFAIR Project (D13.1) Cultural Heritage Mapping Report: Practices and policies supporting Cultural Heritage image sharing platforms,' Zenodo. <https://doi.org/10.5281/zenodo.7659002>
- Koers, H. *et al.* (2020) 'Recommendations for Services in a FAIR Data Ecosystem,' *Patterns* [Preprint]. <https://doi.org/10.1016/j.patter.2020.100058>
- Koster, L. & Woutersen-Windhouwer, S. (2018) 'FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections,' *Code4Lib Journal*, 40. <https://hdl.handle.net/11245.1/4ee73567-5ddd-4bbe-9188-173ad5b0f18b>
- Library of Congress (2015) 'PREMIS Data Dictionary for Preservation Metadata, Version 3.0,' <http://www.loc.gov/standards/premis/v3/>
- Lin, D., Crabtree, J., Dillo, I. *et al.* (2020) 'The TRUST Principles for digital repositories,' *Sci Data* 7, 144. <https://doi.org/10.1038/s41597-020-0486-7>
- Manchester, E.J. (2022a) 'Announcing LC Labs Data Sandbox and 3 New Data Packages,' *Library of Congress Blogs*. <https://blogs.loc.gov/thesignal/2022/12/announcing-lc-labs-data-sandbox-and-3-new-data-packages/>
- Manchester, E.J. (2022b) 'Now Playing: the CCHC Data Jam!' *Library of Congress Blogs*. <https://blogs.loc.gov/thesignal/2022/12/now-playing-the-cchc-data-jam/>
- Metadata Authoring Systems (n.d.) 'ImageSnippets.' <https://imagesnippets.com/index2.html>
- Musen, M.A.. *et al.* (2022) 'Modeling community standards for metadata as templates makes data FAIR,' *Sci Data* 9, 696. <https://doi.org/10.1038/s41597-022-01815-3>

National Digital Stewardship Alliance (2019) 'Levels of Digital Preservation, Version 2.0,' <https://ndsa.org/publications/levels-of-digital-preservation/>

Nicolucci, F., & Di Giorgio, S. (n.d.) *Guidelines to FAIRifying Data: Easy guide for GLAMs and researchers in the Digital Humanities.* https://pro.europeana.eu/files/Europeana_Professional/Event_documentation/Webinars/ECC_Webinar_FAIRifyingData.pdf

OCLC (2018) *Best practices for creating sharable metadata, OCLC Support.* https://help.oclc.org/Metadata_Services/CONTENTdm/Get_started/best_practices

OpenGLAM Working Group (n.d.) 'OpenGLAM Principles,' OpenGLAM. <https://openglam.org/principles/>

Padilla, T., Scates Kettler, H., Shorish, Y., Varner, S. (2023) *Collections as Data – Part to Whole.* <https://collectionsasdata.github.io/part2whole/>

Padilla, T. *et al.* (2018) 'Santa Barbara Statement on Collections as Data – Always Already Computational: Collections as Data,' Zenodo. <https://doi.org/10.5281/zenodo.3066209>

Raemy, J.A., Jung, J. and Kunze, J. (2022) 'Images and the promise of ARKs with IIIF,' *ARK Alliance*, 23 November. <https://arks.org/blog/images-and-the-promise-of-arks-with-iiif/>

'Rights Statements' (n.d.) RightsStatements.org. <https://rightsstatements.org/en/>

Romacker, M. (2022) 'From application-centric to data-centric using FAIR architectures and resources,' EOSC Symposium 2022 [presentation]. <https://symposium22.eoscfuture.eu/symposium/keynotes/>

Social Networks and Archival Context (n.d.) 'About SNAC.' <https://portal.snaccooperative.org/about>

Tasovac, T., Chambers, S. & Tóth-Czifra, E. (8 October 2020) 'Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper,' *HAL*. <https://hal.archives-ouvertes.fr/hal-02961317>

'TK Labels,' Local Contexts, <https://localcontexts.org/labels/traditional-knowledge-labels/>

Tóth-Czifra, E. (2019) 'The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem,' *HAL*. <https://shs.hal.science/halshs-02115505>

Wilkinson, M.D. *et al.* (2016) 'The FAIR Guiding Principles for scientific data management and stewardship,' *Scientific Data*, 3(1), p. 160018. <https://doi.org/10.1038/sdata.2016.18>

Wilkinson, M.D. *et al.* (2019) 'Evaluating FAIR maturity through a scalable, automated, community-governed framework,' *Scientific Data*, 6(1), p. 174. <https://doi.org/10.1038/s41597-019-0184-5>

Appendix A – Recommendations aligned to FAIR Principles

Recommendation 1 Citation Model	Recommendation 2 Transparency	Recommendation 3 Data Documentation	Recommendation 4 Licencing & Rights	Recommendation 5 Delivery
<ul style="list-style-type: none"> • A2: metadata are accessible, even when the data are no longer available • R1.3: (meta)data meet domain-relevant community standards 	<ul style="list-style-type: none"> • F2: data are described with rich metadata • A2: metadata are accessible, even when the data are no longer available • I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation • I2: (meta)data use vocabularies that follow FAIR principles • I3: (meta)data include qualified references to other (meta)data • R1: (meta)data are richly described with a plurality of accurate and relevant attributes • R1.2: (meta)data are associated with detailed provenance • R1.3: (meta)data meet domain-relevant community standards 	<ul style="list-style-type: none"> • F2: data are described with rich metadata • A2: metadata are accessible, even when the data are no longer available • I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation • I2: (meta)data use vocabularies that follow FAIR principles • I3: (meta)data include qualified references to other (meta)data • R1: (meta)data are richly described with a plurality of accurate and relevant attributes • R1.2: (meta)data are associated with detailed provenance • R1.3: (meta)data meet domain-relevant community standards 	<ul style="list-style-type: none"> • F2: data are described with rich metadata • I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation • I2: (meta)data use vocabularies that follow FAIR principles • R1.1: (meta)data are released with a clear and accessible data usage license 	<ul style="list-style-type: none"> • F1: (meta)data are assigned a globally unique and persistent identifier • F3: metadata clearly and explicitly include the identifier of the data it describes • A1: (meta)data are retrievable by their identifier using a standardised communications protocol • A1.1: the protocol is open, free and universally implementable • I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation • I2: (meta)data use vocabularies that follow FAIR principles

Appendix B – Mind map of thematic headings informing the recommendations in accessible list format

Mutability

Data mutability

- Files may be updated to preserve accessibility
- New imaging technology allows for better images
- Image files are often surrogates for real-world originals
 - ◆ Is it all metadata or data? Where do we draw the line?
- Who creates the authoritative image file?

Metadata mutability

- Community-centred description encourages responsive records
- Language matters
 - ◆ Ethics/reparative description
 - ◆ Translation
 - ◆ Plain English <https://www.nala.ie/>
- Controlled vocabularies and authority files get updated
- Metadata may be expressed differently on different platforms (mapping not always 1-to-1)
 - ◆ Is it desirable to harmonise with aggregators?
 - ◆ Do other types of repositories support metadata versioning?

Transparency

- Preservation actions are not visible
- Preservation metadata is not shared
- Reduced granularity of metadata when shared
 - ◆ Where does the most complete record reside?
- Process of selection not always disclosed
- Sometimes 'data' (image files) aren't shared - what is the data?
- Is there a conceptual difference between 2D and 3D images or just a technical one?
- Provenance information is not always exposed
 - ◆ Provenance for physical objects is easier to trace than for digital objects

Technology

- How are search results prioritised?
- URIs support LOD and resource retrieval
- System of identifiers not linked to web technologies (accession number, record number, barcode)
- Is OAI-PMH on its way out?
- Are 3D images able to be managed the same way as 2D?

Rights

- Lack of distinction between copyright and licences
- Implied attribution required (of the repository)
 - ◆ Request to acknowledge needs to be formalised

- Metadata licensing needs to be addressed
- Assumption that metadata is in the public domain
- Recognition of community rights and ownership

Costs

- Can cultural heritage afford DOIs? At what granular level?
 - ◆ PID guide for cultural heritage, see: <https://www.pidwijzer.nl/en>
- There is labour associated with metadata creation that is not currently valued
- Avoiding duplication of effort, both across institutions and researchers
- Privileging of collections in well-funded institutions
- Charging for certain types of use helps to fund collections work (often, literally the digitising requested)

Appendix C – Image Sharing Platforms Summary Table

Image Sharing Platforms - Summary Table from D13.1 with addition of citation field

	DigitalNZ	DPLA	Europeana	Wikimedia Commons	Internet Archive	Flickr
Geo Scope	National	National	European	Global	Global	Global
Repository	Yes*	No	No	Yes	Yes	Yes
Exhibition ^a	Yes	Yes	Yes	No	No	Yes
Aggregator ^b	Yes	Yes	Yes	No	No	No
No. of Images	4.7m	12m	31m	83m	4.5m	5 billion
Open APIs	Yes	Yes	Yes	Yes	Yes	Yes
Viewing version	Access/ Thumbnail	Access/ Thumbnail	Access/ Thumbnail	Full Res	Full Res	Full Res
Export Format	WebP	JFIF	JPEG	JPEG**	JPEG**	JPEG**
EXIF/XMP	No	No	No	Yes	Yes	Yes
Data Model	Supplejack	DPLA-MAP	EDM	Wikibase	Unclear	Unclear
Metadata Standard	DC	DC	DC	DC	DC	Unclear
LOD support ^c	Yes	Yes	Yes	Yes	No	No
IIIF ^d	Unclear	Yes	Yes	Yes	Yes	No
Creative Commons ^e	Yes	Yes	Yes	Yes	Yes	Yes
Cite Option	No	Yes	No	No	No	No
^a Exhibition The platform provides additional services such as curated exhibitions ^b Aggregator Platform that links back to content providers' site for full image and metadata ^c LOD support Linked Open Data supported/on roadmap, but not necessarily implemented ^d IIIF support IIIF is supported but not necessarily implemented for all content providers ^e Creative Commons CC defines standard licence information for content reuse * DigitalNZ provides a repository, limited to at-risk content from small collections ** JPEG is the predominant file format, however where content has been uploaded as TIFF, PNG, GIF, SVG it may also be downloaded as same						



DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

