



## WorldFAIR

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	<a href="http://worldfair-project.eu">http://worldfair-project.eu</a>

# D13.1 Practices and policies supporting Cultural Heritage image sharing platforms

Work Package	WP13 - Cultural Heritage
Lead Author (Org)	Beth Knazook (DRI)
Contributing Author(s) (Org)	Joan Murphy (DRI)
Due Date	28.02.2023
Date	21.02.2023
Version	1.0 <b>DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION</b>

DOI

<https://doi.org/10.5281/zenodo.7659001>

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	02.02.2023	Beth Knazook, Joan Murphy	Draft for internal review
1.0	21.02.2023	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
DC	Dublin Core
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
HEI	Higher Education Institution
ICA	International Council on Archives

ICOM	International Council of Museums
LOD	Linked Open Data
MARC	MAchine Readable Cataloging
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
PID	Persistent Identifier
SAA	Society of American Archivists
URI	Uniform Resource Identifier
XML	eXtensible Markup Language

### **A note on hyperlinks**

---

This report contains many embedded links to documents, websites and other resources. A complete list of links is available at the end of the bibliography.

## Executive Summary

Deliverable 13.1 for the WorldFAIR Project's Cultural Heritage Work Package (WP13) outlines current practices guiding online digital image sharing by institutions charged with providing care and access to cultural memory, in order to identify how these practices may be adapted to promote and support the FAIR Principles for data sharing.

The report has been compiled by the Digital Repository of Ireland as a key information resource for developing the recommendations forthcoming in Deliverable 13.2. The DRI is Ireland's national repository for the arts, humanities and social sciences. A Working Group of cultural heritage professionals has been invited to contribute feedback.

There are well-established standards and traditions driving the various approaches to image sharing in the sector, both local and global, which influence everything from the creation of digital image files, their intellectual organisation and level of description, to statements of rights governing use. Additionally, there are technological supports and infrastructures that have emerged to facilitate these practices which have significant investment and robust community support. These practices and technologies serve the existing communities of users well, primarily the needs of government, business and higher education, as well as the broader general public. Recommendations for adapting established collections delivery mechanisms to facilitate the use of cultural heritage images as research data would ideally not supersede or duplicate processes that also serve these other communities of users, and any solutions proposed in the context of the WorldFAIR Project must be made in respect of these wider contexts for image sharing.

## Table of contents

<b>Executive Summary</b>	<b>4</b>
<b>1. Introduction</b>	<b>7</b>
1.1 Background	7
1.2 Definitions	11
<b>2. Image interoperability</b>	<b>11</b>
2.1 Enabling practices	12
2.1.1 Policies	12
2.1.2 Metadata standards	20
2.1.2.1 Metadata structure standards	20
2.1.2.2 Descriptive metadata and data value standards	22
2.1.3 Data formats	24
2.1.4 Rights and licences	26
2.2 Enabling technologies	27
2.2.1 Web technologies – HTTP, HTML, XML	28
2.2.2 Metadata harvesters – Z39.50, OAI-PMH	28
2.2.3 Semantic web technologies	29
2.2.4 International Image Interoperability Framework (IIIF)	30
2.2.5 Embedded metadata	31
2.3 Image sharing platforms	31
2.3.1 DigitalNZ	34
2.3.2 DPLA	36
2.3.3 Europeana	38
2.3.4 Wikimedia Commons	40
2.3.5 Internet Archive	42

2.3.6 Flickr	44
<b>3. Summary of key issues raised for the WP13 case study</b>	<b>47</b>
3.1 Positioning DRI in the landscape	47
3.2 Conclusions and next steps	49
<b>4. Bibliography</b>	<b>51</b>
4.1 List of web links (in alphabetical order)	53

## 1. Introduction

A key output of the WorldFAIR Project will be the production of a Cross-Domain Interoperability Framework (CDIF) that facilitates the sharing of research data across previously siloed domains of research practice, guided by expert assessments of how the various domain activities relate to the FAIR principles.<sup>1</sup> In order to support the creation of that framework, each Work Package is tasked with the creation of a FAIR Implementation Profile (FIP) for their domain or subdomain of work in addition to other activities leading to better FAIR alignment. This report recognises that for Cultural Heritage to be represented as a research data-producing domain within the CDIF, and FAIR-aligning practices proposed, then some sense of the overarching approaches and challenges to data interoperability within the sector must first be assessed.

The purpose of the Cultural Heritage Work Package Deliverable D13.1 is to identify how image sharing platforms in the cultural heritage landscape already facilitate the interoperability of both image data and associated metadata. While the overall outcome of the case study will be the production of a model for implementing FAIR recommendations at the Digital Repository of Ireland (DRI), it is also an opportunity to highlight achievements in the sector which may usefully inform work in other domains. In that respect, this report is both a valuable opportunity to review and situate the DRI's position as an image-sharing platform in the cultural heritage landscape, as well as a chance to raise questions about how FAIR may be perceived in that landscape.

This report starts with an overview of the common threads in professional practice in the diverse collecting fields that engage with cultural heritage objects, recognising that a complete picture of all practices is not possible and acknowledging there will be gaps in information around local implementations. In order to define common practice, research for this report primarily focused on the websites of professional bodies, standards organisations and national institutions responsible for setting trends within the sector. This is followed by a detailed look at six examples of platforms and the technologies that facilitate and further shape the way that cultural heritage images are shared. The conclusions examine DRI's position in this landscape.

### 1.1 Background

The FAIR principles were published in 2016 to provide a baseline strategy for all domains of research to make their data broadly reusable by others. They describe approaches to findability, accessibility, interoperability and reproducibility for both data and accompanying metadata and although they are not prescriptive in what may be considered FAIR, the metrics which have been derived to assess FAIRness have largely focused on a limited number of machine actionable criteria. These criteria essentially ask: Can the research data and metadata be easily discovered and shared between

---

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

systems, and manipulated by technologies, while also maintaining their integrity and usefulness as authoritative resources?

These are questions that cultural heritage institutions have been grappling with for far longer, and it is arguably the success of the sector's efforts to create discoverable, understandable and machine actionable metadata that have made the complexity of the work somewhat invisible to the average user. Online Public Access Catalogues (or OPACs) have been at the forefront of library work since the 1960s, followed by the wide adoption of the MARC (MACHine Readable Cataloguing) metadata standard over the course of the 1970s and 1980s, which remains in popular use today. This information is exchanged between library systems and sometimes aggregated through union catalogues to improve discoverability, facilitate interlibrary loan and drive acquisition decisions (e.g. [WorldCat](#)). Although much has changed in the information landscape since the dawn of the World Wide Web, the organising principles guiding how information is made intelligible and accessible remain largely the same.

According to *The MARC 21 Formats: Background and Principles* (1996), a MARC record involves three elements: record structure, content designation, and the content itself.<sup>2</sup> In other words, there is an organising structure ensuring that records are well-formed, the content within the record is characterised and explicitly identified so that it can be manipulated, and the required content is appropriately provided (in the case of MARC and many other data structure standards, the expression of the content is also governed using related description and data value standards).

It is important to note that the MARC standard, like other metadata standards in the cultural heritage domain, are based on much older analogue systems (see Figure 1) that were primarily focused on metadata exchange *only*. The data, for the most part, remained physical and therefore physically separated from the information systems that allowed access. It was necessary within the analogue model to explain everything one might need to know about an object before sending the user out to retrieve the object, and metadata records created today still prioritise the transcriptions of titles, signatures or other markings that may be perfectly visible in the digital image file (and possibly even captured by OCR), as well as describing what kind of physical format the object takes and providing crucial unique identifiers and location information that allows for materials to be retrieved.

---

<sup>2</sup> American Library Association, MARBI Committee (Revised 1996) *The MARC 21 Formats: Background and Principles*, <https://www.loc.gov/marc/96principl.html>.



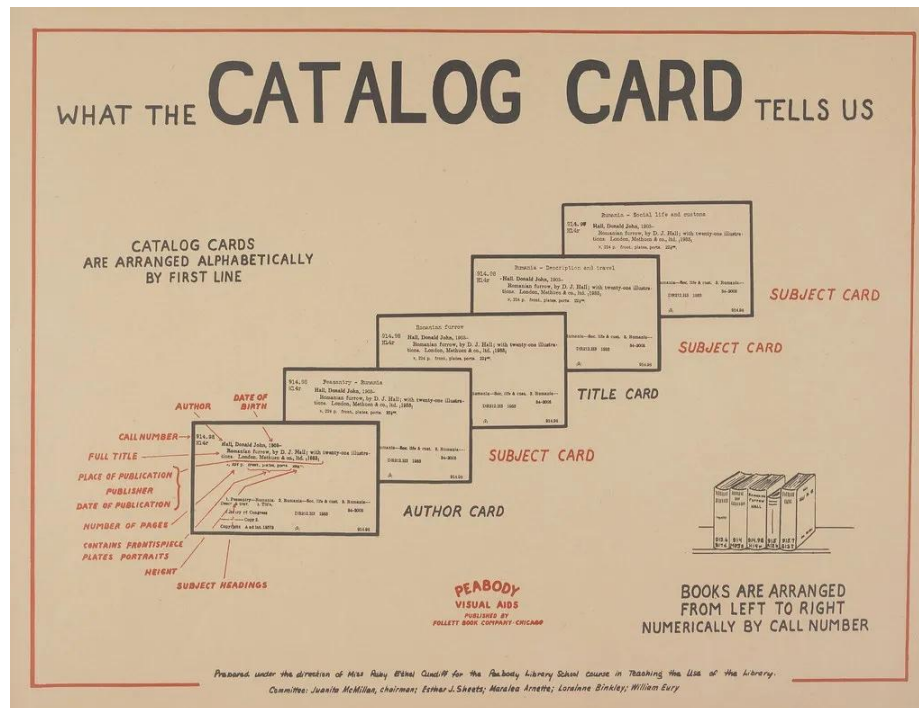


Figure 1 - 'What the Catalog Card Tells Us.' Source: Char Booth via flickr.com, CC BY-NC-SA 2.0

Some of the image sharing platforms reviewed in this report (e.g. Europeana<sup>3</sup> and DPLA<sup>4</sup>) are fundamentally aggregators of metadata only. Even though thumbnail or low resolution reference images may be made available to the end user, it is reasonable to say that these are not considered to be original or necessarily authoritative representations, but attempts to augment the informational content of the metadata record.<sup>5</sup> The terminology used consistently across the sector refers to the digital content shared in online collections as 'access images' or 'surrogates' whether derivatives of born-digital content (images created from digital devices) or digitised images (images made from analogue or physical objects). In these cases, the metadata records will usually link back to metadata records at the originating institutions that contributed the content. By following those links, the end user is able to download or request a high resolution image and view additional contextual information, which may be presented according to a different metadata standard and/or with locally defined fields not available through the aggregator.

<sup>3</sup> <https://www.europeana.eu/en>

<sup>4</sup> The Digital Public Library of America: <https://dp.la/about>

<sup>5</sup> This is highly dependent on the skill and resources of the institution doing the digitising. Well-resourced organisations have much greater capacity to ensure the quality of the digital representation, while others may not necessarily be able to warrant that. A 2007 OCLC Report which still resonates today argued 'Scaling up digitization of special collections (here defined as non-book collections, such as photographs, manuscripts, pamphlets, minerals, insects, or maps) will compel us to temper our historical emphasis on quality with the recognition that large quantities of digitized special collections materials will better serve our users.' Ricky Erway and Jennifer Schaffner (2007) 'Shifting Gears: Gearing Up to Get Into the Flow,' OCLC, <https://www.oclc.org/content/dam/research/publications/library/2007/2007-02.pdf>.

The institutional context for the collection may be valuable information in its own right and important to communicate alongside the record, as the organisations themselves influence how materials are described and how they might be perceived. The cultural heritage sector is divided into four primary areas of work: Galleries, Libraries, Archives and Museums (often referred to collectively by the acronym ‘GLAM’). Each of these areas of professional practice are separately guided by local, national and international professional societies, academic training programs, standards and codes of practice, which frames and shapes the nuanced approaches to collections delivery within the space. As digital collections have become at once a more intense demand on resources and an important means of reaching audiences, the persistence of siloed approaches to cultural heritage have for some time prompted concerned voices to raise calls for better collaboration.<sup>6,7</sup>

In response, significant work has been done already to support interoperability within the cultural heritage sector, particularly around the standardisation and exchange of metadata. In addition to the authoritative *Getty Metadata Standards Crosswalk*, which provides an intellectual mapping of over 15 standards, there are crosswalks that support contributions to individual image-sharing platforms as well as crosswalks which directly map one standard to another (for example, [Dublin Core to MARC](#)).<sup>8</sup> Many metadata schemas and controlled vocabularies have also been individually mapped to the [Resource Description Framework](#) (RDF), a standard model for data interchange on the Web, or they may be expressed as [XML](#) or [JSON](#). The fields (or terms) in these metadata standards are also uniquely identified with [Uniform Resource Identifiers](#) (URIs) to support their use as [Linked Open Data](#) (LOD).

Collaborative platforms have emerged to further these connections as well as fill a resource gap around money, skills, and technological expertise for many institutions operating in this sector. (The calls for collaboration repeating throughout the last two decades have been as much focused on improving capacity as providing technical solutions.) Some of these platforms, such as [Wikimedia](#), [Flickr](#) or [Internet Archive](#), do not have particular professional alignments or an associated ‘duty of care’ for cultural heritage materials yet have emerged as reliable sources of data and metadata from the cultural heritage sector. This report explores the policies, standards, technologies and practices that influence those platforms.

---

<sup>6</sup> Diane Zorich, Günter Waibel, and Ricky Erway (2008) ‘Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums.’ OCLC Research, <https://doi.org/10.25333/X187-3W53>.

<sup>7</sup> Guy Berthiaume, ‘If You Want to Go Far, Go Together: The Collaboration among the GLAM Community in Canada (2016–2019).’ *Research Library Issues*, no.300 (2020): 6–17, <https://publications.arl.org/rli300/1>.

<sup>8</sup> Patricia Harpring (revised 25 January 2022) ‘Metadata Standards Crosswalk,’ Getty Research Institute, [https://www.getty.edu/research/publications/electronic\\_publications/intrometadata/crosswalks.html](https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html).

## 1.2 Definitions

For the purpose of this report, the following definitions have informed the usage of particular terminology:

*Aggregator:* A platform or site that provides links back to the site of the original content provider.

*Common Practice:* Tools, workflows, metadata standards, or policies that are widely used to manage GLAM collections regardless of whether legally mandated or regulated.

*Cultural Heritage Sector:* Institutions, organisations and professionals tasked with the care, management, interpretation, dissemination and preservation of cultural memory. The sector may include publicly or privately funded institutions, non-profits, corporations, research and higher education institutions, as well as volunteer and community-led organisations. The European Commission defines a cultural heritage institution as ‘a publicly accessible library or museum, an archive or a film or audio heritage institution,’ as defined in Article 2, point (3) of Directive (EU) 2019/790.<sup>9</sup> This term may be used interchangeably with ‘GLAM sector’ throughout this report.

*Digital Preservation:* A series of ‘managed activities necessary to ensure continued access to digital materials for as long as necessary’.<sup>10</sup>

*End User:* People using image sharing platforms or online collections databases to obtain images and associated metadata for any purpose; the general public.

*Images:* Digital raster images, created through digitisation as a means of representing a real-world physical object, or images created via digital technologies as digital originals (often referred to as ‘born-digital’ content). Images are assumed to be 2D unless explicitly stated otherwise (e.g. 3D).

*Image Sharing:* Making image files and their associated metadata publicly available and reusable by an external audience.

*Metadata:* Information which describes aspects of a resource, literally ‘data about data.’

## 2. Image interoperability

As explored in section 1 above, the cultural heritage sector has been actively engaged in the sharing of images and associated metadata with researchers and the general public for a long time, even

---

<sup>9</sup> European Commission (2019) ‘DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC,’ *Official Journal of the European Union*, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>

<sup>10</sup> Neil Beagrie and William Kilbride, eds. (2015) *Digital Preservation Coalition Handbook (2nd Revised Edition)*, <https://www.dpconline.org/handbook/>.

though cross-sectoral interoperability has been a comparatively recent development. The purpose of this section of the report is to highlight the policies, technologies and standards that influence how image sharing platforms leverage that expertise to enable the sector to deliver images to end users. The following is not meant to be a representation of the whole scope of practices and services across the GLAMs (in contrast to the FIPs being created through this case study and across the WorldFAIR project, which provide that additional case-by-case specificity). Here we consider the factors that support interoperability under three broad headings:

- Enabling practices, such as Policies, Metadata, Rights and Licensing
- Enabling technologies, such as OAI-PMH and IIIF
- Examples of cross-sectoral image sharing platforms, such as DigitalNZ and Europeana

## 2.1 Enabling practices

Image sharing platforms which serve the cultural heritage sector are underpinned by the policies, systems and standards in use across the GLAMs. They utilise the metadata standards developed by the communities they serve and they support exchange using community-approved technologies and systems, typically selected from open source and non proprietary solutions. This section will provide a high-level overview of those practices endorsed by international professional associations and leading institutions which determine the selection, documentation and technical delivery of cultural heritage image data and metadata.

### 2.1.1 Policies

Each of the image sharing platforms reviewed below have policies that inform the use of the sites, such as Terms of Use, Privacy Policies and Depositor's Agreements. The professional associations that support the GLAMs also create or share key guidance documents and standards of practice with their communities that directly influence both the policies of the image sharing platforms and the type of content that is published on them. These documents empower institutions to select and represent collections according to available budgets, needs, and their varying levels of capability to sustain the digital files and records. Standards of practice further promote discoverability and accessibility through the timely processing of collections, the use of unique identifiers to aid the retrieval of materials, the creation of publicly accessible discovery aids and the inclusion of clear copyright information to allow the end user to determine whether their intended reuse of any collections material is allowed. Metadata records are always made publicly available regardless of whether any part of the collection is restricted from access, although processing backlogs created by lack of resourcing and the volume of materials sometimes see these efforts unfortunately delayed for a considerable length of time.

Digitisation and digital preservation practices are further managed by a suite of policies, from those guiding collection acquisition and disposition, processing and digital imaging, to those that address preservation actions that ensure the ongoing accessibility and reuse of items through transformations to new formats. The latter ensures that the files made available to end users

remain appropriate for access via current technologies, while the complementary policies ensure that there is transparency around what is collected, how long it is kept for, and what activities surround the appropriate disposal should materials need to be removed from the collection. In the case of archival, museum or special collections, this may involve a requirement to find another appropriate stewarding organisation to transfer the collections or a public announcement of the intent to offer collections for sale. Many of the activities required to create digital files and keep them accessible through image sharing platforms will likely be performed by the contributing institution as part of the stewardship obligation outlined in these policies. There is a reasonable expectation that the vast majority of materials held by GLAMs will be preserved and remain available indefinitely.

That sense of permanence does not necessarily extend to the image sharing platforms that help to expose digital content. These platforms typically allow contributing institutions to maintain and update both data and metadata periodically, but records of these updates and version access to earlier iterations of the data are often not made publicly available. Existing files may be updated and replaced, new images taken and the old ones deleted or digital files may be added to records that previously had no associated image. Long-term file preservation, which includes migration of file formats to avoid obsolescence, is generally not offered as a service although there is clearly some effort to ensure persistence of the files uploaded to the platform.

The mutability of cultural heritage records is not only an important factor for accessibility but also integral to the institution's ability to maintain relevance and respect for the communities they serve. As the keepers of authoritative sources of cultural memory, GLAM professionals are aware that what they choose to collect and how they represent those collections online will influence research agendas, community perspectives and the ongoing cultural value of the collections.<sup>11</sup> For this reason, selection criteria for digitisation projects often consider not only an institution's legal right to share images, but its moral obligations and capacity to sustain the collections.

Policy guidance and reference documents from the GLAMs do not address metadata preservation separately from digital object preservation. The [PREMIS](#) standard recommends that descriptive information is retained alongside technical metadata and preservation event information, particularly around ownership and rights, but PREMIS is not intended to be user-facing metadata. Metadata records that accompany materials in online platforms may be removed from public display and/or deleted from online aggregator platforms for a number of reasons, including deaccession of the original items from institutional collections or at the end of a long-term loan agreement.<sup>12</sup> Internal acquisition records or accession records would ensure this information is kept

---

<sup>11</sup> Merrilee Proffitt (2008) 'Impact of Digitization on Scholarship and Collecting,' *Hanging Together*, the OCLC Research Blog, <https://hangingtogether.org/impact-of-digitization-on-scholarship-and-collecting/>.

<sup>12</sup> Though long-term loan is often discouraged, it does occasionally happen that the stewardship of collections is assigned to a third-party organisation. See, for example: Teijamari Jyrkkiö, ed. (2010) 'Long-term loans best practices report,' Finnish National Gallery, [https://www.ne-mo.org/fileadmin/Dateien/public/topics/Collection\\_Mobility/Long-term\\_loans\\_best\\_practices\\_report.pdf](https://www.ne-mo.org/fileadmin/Dateien/public/topics/Collection_Mobility/Long-term_loans_best_practices_report.pdf).

by the institution but not be made publicly available. It is unclear how provenance information would be communicated if metadata records were transferred between institutions as part of a collections transfer.

The following table further summarises professional practice recommendations that shape key areas of activity also addressed by the FAIR principles, namely: the acquisition and selection of collections, processing and delivery decisions, approaches to description, level of openness, digitisation standards and preservation.

Table 1 - Professional Practice Guidance and Standards

	Acquisition and Selection for Digitisation	Discovery	Description	Level of Openness	Digitisation Technical Standards	Preservation
	Digital images should be acquired in accordance with the institution's collections policy and in consideration of national and international provisions, including the <a href="#">United Nations Declaration on the Rights of Indigenous Peoples</a> . The collection should be under intellectual control (accessioned and processed) before being digitised. It is generally recommended not to duplicate the efforts of other institutions. The	Emphasis on the timely preparation of collection information and public dissemination through discovery tools such as finding aids, registers and catalogues. Focus on necessary minimums for delivery. Records may be enhanced at future dates, as needed or as time and resources allow. Accession or record numbers provided for retrieval, though these numbers may be applied to aggregations rather than individual objects.	Multi-level description focuses on a small set of characteristics describing aggregations of records, and information is not repeated at lower levels. Archival description is always contextualised through provenance. Documentation of archaeological, cultural and artistic objects should include information necessary to identify them in case of loss or theft. Responsive and reparative description may require additions to the interpretive frameworks and access pathways to records. Wide uptake of standards but there are	All collections should be described and their metadata made publicly available, even in cases where the material itself may be restricted from access. Some metadata may be restricted where privacy laws apply. Copyright permission for digitisation and image sharing is typically requested at the time of acquisition, but this permission does not include reusers of the digital collections. Access to material is not limited to any particular audience. Partnerships	Digitisation capture standards recommend the minimum resolutions sufficient to capture all significant details in the source document (varies based on source). Decisions relating to bit depth and scale of image capture (cropped or complete object, image side only or front and back, 3D capture) are decided independently based on whether the physical format has an added informational value. Avoid digitising from intermediates if originals are available and in reasonable	Commitment to continued access to content in the care of the institution. Recommendations for reappraisal and deaccession indicate that each step of the process must be documented and records retained. There are legal obligations to maintain certain types of records. Highest level of digital preservation requires monitoring of content, regular fixity checking of assets and the ability to replace deteriorated or obsolete files where necessary. Preservation choices may take into

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

	institution must ensure it has the appropriate legal and ethical permissions for any images selected for sharing, and community consultation is often recommended. The institution should have the capacity to sustain the collections selected for digitisation.	Unique filenames also used as identifiers.	institutions that have devised local metadata. The use of data value standards or conventions for naming are encouraged to support precision in searching and browsing collections.	with outside vendors should not substantially restrict the open sharing of digital images and metadata.	condition.	consideration the informational value of the format as well as the content.
<b>Examples</b>						
Society of American Archivists (SAA)	<a href="#">Brooklyn Historical Society Photographic Digitization and Cataloging Manual</a>  <a href="#">Protocols for Native American Archival Materials</a>	<a href="#">More Product, Less Process: Revamping Traditional Archival Processing</a>  <a href="#">ACRL/RBMS-SAA Guidelines on Access to Research Materials in Archives and Special Collections Libraries</a>	<a href="#">EAD: Encoded Archival Description</a>  <a href="#">Describing Archives: A Content Standard</a>  <a href="#">Archives for Black Lives in Philadelphia</a>	<a href="#">ACRL/RBMS-SAA Guidelines on Access to Research Materials in Archives and Special Collections Libraries (2020)   Society of American Archivists</a>  <a href="#">Well-intentioned Practice for Putting</a>	<a href="#">Brooklyn Historical Society Photographic Digitization and Cataloging Manual</a>  <a href="#">Principles to Guide Vendor/Publisher Relations in Large-Scale Digitization Projects of Special Collections</a>	<a href="#">SAA Museum Archives Section Working Group Example</a>  <a href="#">Guidelines for Reappraisal and Deaccessioning</a>  <a href="#">Levels of Digital Preservation</a>





		<a href="#">(2020)   Society of American Archivists</a> <a href="#">Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)</a> <a href="#">Open Archives Initiative Object Re-use and Exchange (OAI-ORE)</a>		<a href="#">Digitized Collections of Unpublished Materials Online</a> <a href="#">Orphan Works: Statement of Best Practices</a>	<a href="#">Materials</a>	<a href="#">PREMIS Data Dictionary for Preservation Metadata, Version 3.0 (Library of Congress)</a> <a href="#">ISO16363:2012</a>
International Council on Archives (ICA)	<a href="#">Digital Records Curation Programme</a> <a href="#">The UNESCO/PERSIST Guidelines for the selection of digital heritage for long-term preservation</a>	<a href="#">Principles of Access to Archives</a>	<a href="#">Archival Arrangement and Description</a>	<a href="#">ICA Copyright Declaration</a> <a href="#">Reference Dossier on Archival Claims</a> <a href="#">Technical Guidance for Managing Archives with Restrictions</a>	<a href="#">Preservation issues in digitizing historical photographs</a> <a href="#">Digital imaging for photographic collections</a>	<a href="#">Preserving archives</a> <a href="#">The UNESCO/PERSIST Guidelines for the selection of digital heritage for long-term preservation</a>
International Federation of Library Associations and Institutions (IFLA)	<a href="#">IFLA Repository: The UNESCO/PERSIST Guidelines for the Selection of Digital Heritage for Long-Term Preservation - 2nd Edition</a>	<a href="#">BIBFRAME s Application Method for Reflecting LRM in Linked Data</a> <a href="#">Cataloger acceptance and use of semiautomated subject</a>	<a href="#">MARC 21 Format for Bibliographic Data: Table of Contents (Network Development and MARC Standards Office, Library of Congress)</a>	<a href="#">IFLA Statement on Open Library Data</a> <a href="#">Navigating Copyright for Libraries – Purpose and Scope</a>	<a href="#">Guidelines for Digitization Projects for collections and holdings in the public domain</a> <a href="#">Minimum Digitization Capture</a>	<a href="#">The UNESCO/PERSIST Guidelines for the Selection of Digital Heritage for Long-Term Preservation</a> <a href="#">IFLA Principles for the</a>

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

		<a href="#">recommendations for web scale linked data systems</a>	<a href="#">New horizons: emerging metadata standards and practices in the 21st century</a>		<a href="#">Recommendations</a>	<a href="#">Care and Handling of Library Materials</a>
Online Computer Library Center (OCLC)	<a href="#">Retrospective Accessioning</a>  <a href="#">Planning an Imaging Project</a>	<a href="#">Shifting Gears: Gearing Up to Get Into the Flow</a>  <a href="#">Linked Data Research</a>  <a href="#">IIIF: Improving the Interoperability of Image Materials</a>  <a href="#">Library Collaboration as a Strategic Choice</a>	<a href="#">Library Linked Data in the Cloud, Chapter 2: Modeling Library Authority Files</a>	<a href="#">Introduce Balance in Rights Management</a>  <a href="#">Well-intentioned practice for putting digitized collections of unpublished materials online</a>  <a href="#">Good Terms – Improving Commercial-Noncommercial Partnerships for Mass Digitization</a>	<a href="#">Technical Standards for Digital Conversion of Text and Graphic Materials</a>  <a href="#">Streamlining Photography and Scanning</a>  <a href="#">Planning an Imaging Project</a>	<a href="#">Preservation metadata for digital materials</a>





WorldFAIR

DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

International Council of Museums (ICOM)	<a href="#">Standards on Accessioning of the International Council of Museums</a> <a href="#">Encouraging Collections Mobility   ICOM UK</a>	<a href="#">Key Concepts of Museology</a>	<a href="#">Object ID - International Council of Museums</a>	<a href="#">Statement on the independence of museums</a> <a href="#">Museums and inclusion</a> <a href="#">FAQs Author's Right, Copyright and Free Licences for Culture on the Web</a>		<a href="#">Sustainability</a>
---	---	---	--	--	--	--------------------------------



## 2.1.2 Metadata standards

Although archives, libraries and museums often overlap considerably in the types of materials they collect and make available, they organise information about their collections according to substantially different conceptual models and metadata standards. Not only may the same object be described in different ways, using different required metadata field elements, but depending on the domain of knowledge that produced the description each object might also have significantly different relationships to other materials.

### 2.1.2.1 Metadata structure standards

Metadata standards describe common ways to organise and communicate various types of information. A structure standard specifically dictates the categories of information (often expressed as ‘fields’ or ‘elements’) that may be used to create a particular type of metadata record.

There are certain elements that are common across the standards in use in cultural heritage, as well as a number of similar elements, which allow for easy mapping from one field in one standard to a related field in another. The *Getty Metadata Standards Crosswalk* highlights the commonality of fields such as: title, description, subject, rights and format.<sup>13</sup> However, a glance at the Getty mapping also quickly reveals the variety and complexity of information expressed by those fields. Some of the common descriptive elements might be represented by multiple fields in one system or a single field in another. While acknowledging this complexity, metadata crosswalks generally only advise which fields *may* be mapped and not *how* to actually implement the mapping. For instance, in the case of a many-to-one relationship where there are several fields that could be mapped to a single field, it is unclear whether the single field should be repeated and the information stored in separate instances of the field or whether the information should in fact be combined into a single instance of the field. This appears to largely be the individual decision of the organisation providing the metadata or providing the system into which the metadata needs to be mapped. Although Online Computer Library Center (OCLC) recommends in its *Best Practices for Creating Sharable Metadata* that ‘when multiple values are needed, the metadata element should be repeated,’ there is also a caveat added that this option might not work well for all [OAI-PMH](#) (Open Archives Initiative Protocol for Metadata Harvesting) harvesters, such that ‘in some cases multiple values (separated by a semicolon) are preferred for accuracy depending upon the level of complexity in configuring a collection using your digital collections management software and the OAI harvesting tool.’<sup>14</sup> Some image sharing platforms provide metadata crosswalks to their depositors, but it is always ultimately the depositor that makes the call as to how to populate the information in the fields from the metadata in their original collections records.

---

<sup>13</sup> Patricia Harpring, ‘Metadata Standards Crosswalk.’

<sup>14</sup> OCLC Support (2018) ‘Best Practices for Creating Sharable Metadata,’  
[https://help.oclc.org/Metadata\\_Services/CONTENTdm/Get\\_started/best\\_practices](https://help.oclc.org/Metadata_Services/CONTENTdm/Get_started/best_practices).

Metadata crosswalks often only map field information (or elements) and not the entire, complex systems governing the use of those elements. For example, the [Functional Requirements for Bibliographic Records](#) (FRBR) conceptual model used in libraries creates relationships between library objects via entity-relationships, expressed through a hierarchical arrangement of Work, Expression, Manifestation and Item. This allows an end user of a library catalogue to search for a copy of Shakespeare's *Hamlet* and find various editions of the play text alongside recordings of theatrical productions, audio commentary, adaptations etc., held at the same or different libraries, because the items are all manifestations of the same work, *Hamlet*.

The [Metadata Encoding and Transmission Standard](#) (METS) is 'a metadata standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library' widely used to document and correctly display digitised pages in books, newspapers, or other multi-page textual documents. METS can be used as a format for exchanging digital library objects (or just their metadata) between different repositories. When paired with [ALTO](#), these XML-based standards can also encode information about the layout, style (including fonts and paragraph breaks) and physical description of the real-world items being digitised, making METS/ALTO extremely popular for digitised historical newspapers. The success of METS in storing and documenting the order of complex structures has made it the backbone of many digital preservation systems as well.<sup>15</sup>

In a similar vein, the Visual Resources Association's [VRA Core](#) metadata structure standard, which was designed to manage art image collections in museums and visual resource libraries, supports part-to-whole relationships between works, part-works and details. This conceptual model might be best explained by using the example of a mediaeval three-panel altarpiece. The altarpiece could be described by a single Work Record in VRA Core, while each of the individual panels may be described by an Image Record with a relationship to the Work. In this case, information that is common to the Work and already expressed in the Work Record is not repeated in the Image Record; the latter only contains information that is particular or unique to that part of the Work.

Archival conceptual models similarly utilise a hierarchical arrangement, but one based on the relationship of works to a person or organising entity. Using the first example above, a copy of *Hamlet* in the records of one individual and another copy of the text in another record group would be represented as discrete and unrelated objects. Connections are not encouraged between items that belonged to different people, although increasing use of subject indexing in archives is making it more possible to browse across archives and discover these kinds of connections.

Efforts towards metadata interoperability that have been made by many of the image sharing platforms examined in this report have largely sidestepped the issue of how the material may be arranged differently in the databases of the GLAM institutions that originated the records (into collections or record groups, series, files or items, as parts of items, derivatives or related works) by creating databases that support 1-to-1 relationships between the data (the image file) and its

---

<sup>15</sup> Eld Zierau (2013) 'PREMIS, METS and preservation metadata: emerging trends and future directions,' Digital Preservation Coalition, <https://www.dpconline.org/docs/miscellaneous/events/856-premismets-zierau-1-1/file>.

metadata (how that file is described and managed). Each digital file is given a unique identifier and is described independently from other images, regardless of whether the image content might be found elsewhere. Metadata structures are mapped almost exclusively to data models based on Dublin Core's flexible and extensible, non-hierarchical layout. In general, image descriptions are not reliant on related collection-level descriptions for any required interpretive or contextual information, although these relationships may be revealed if users click through when links are provided back to an institution's originating records.

#### *2.1.2.2 Descriptive metadata and data value standards*

Descriptive metadata standards are usually developed to be used in conjunction with a metadata structure standard, which means that certain pairings are common (e.g. [EAD](#) with [DACS](#), MARC with [RDA](#), and VRA Core with [CCO](#)) [see Table 2 below]. A wildcard thrown into this system is the [Dublin Core](#) standard, which is often used as both a descriptive and structural standard for image collections, or the structure into which many descriptive standards may be mapped.

It is possible to use a descriptive metadata standard without fully deploying all the fields or elements available in the standard, with some standards having very low or no minimum compliance requirements. For instance, using only one field element from Dublin Core could indicate compliance with the standard even though the 15 core elements are generally recommended. Most descriptive standards have a minimum requirement of at least 2 or 3 fields, but little else dictates how fully the standard is utilised. Even within the same institution, there can be an uneven use of metadata elements across collections, depending on individual choices made during arrangement, processing, digitising and description. Most image platforms allow for records to be edited, which also means that additional metadata elements can potentially be added at any time.

Data value standards (such as controlled vocabularies or authority lists) are often recommended for certain field entries in structural and descriptive metadata schemas when specificity is desired. Where narrative descriptions provide the end user with a nuanced and subjective interpretation of the material, controlled terms provide objectivity in the form of industry-specific or formally accepted terms. For example, the VRA Core 4.0 recommends both the [Union List of Artists Names](#) (ULAN) and the [Library of Congress Name Authority File](#) (LCNAF) as controlled sources for the 'Agent' element. In order to further enhance the precision of these tools, most of the commonly used vocabularies have been mapped for use as Linked Open Data.

Despite the promise of the data value standard, there is a significant margin for error in how these resources are used in metadata records across the GLAMs. Different controlled vocabularies may use identical or near identical terms to represent the same concepts, but vocabularies developed for different purposes will have a broader or narrower degree of specificity in certain areas. In addition, some branches of a vocabulary may be intended to describe different concepts, such as author role, subject or format. As an example, the [Thesaurus for Graphic Materials](#) is technically two thesauri, the TGM I and TGM II, which describe subjects and genre/format terms respectively. The

two thesauri were merged in 2007 but a note field describing the ‘term type category’ still indicates their suitability for metadata fields describing either subject or physical description. Finally, certain vocabulary resources might also allow the cataloguer to construct unique strings from the terms available, such as the [Library of Congress Subject Headings](#), which supports some combinations of established headings and free-floating subdivisions. Using data value standards effectively requires understanding both the scope and structure of the vocabulary as well as the requirements of the metadata fields that accept the vocabulary source. Errors may be exacerbated in image sharing platforms where many contributors using these vocabularies have different needs and levels of understanding.

**Table 2 - Summary of Common Cultural Heritage Metadata Standards**

Acronym	Name	Scope	Type
AAT	<a href="#">Art &amp; Architecture Thesaurus</a>	Global	Data Value Standard
CCO	<a href="#">Cataloging Cultural Objects</a>	Global	Descriptive Standard
CDWA	<a href="#">Categories for the Description of Artworks</a>	Global	Structure Standard
CIDOC-CRM	<a href="#">International Council of Museums International Committee for Documentation Conceptual Reference Model</a>	Global	Data Model
CIDOC-CRMdig	<a href="#">CIDOC Digitisation Provenance Model</a>	Global	Data Model
CONA	<a href="#">Cultural Objects Name Authority</a>	Global	Data Value Standard
DC	<a href="#">Dublin Core</a>	Global	Structure Standard
DACS	<a href="#">Describing Archives: A Content Standard</a>	National	Descriptive Standard
DPLA-MAP	<a href="#">DPLA Metadata Application Profile</a>	National	Data Model
EAD	<a href="#">Encoded Archival Description</a>	Global	Structure Standard
EDM	<a href="#">Europeana Data Model</a>	European	Data Model
FAST	<a href="#">Faceted Application of Subject Terminology</a>	Global	Data Value Standard
FRBR	<a href="#">Functional Requirements for Bibliographic Records</a>	Global	Data Model
ISAD(G)	<a href="#">General International Standard Archival Description</a>	Global	Structure Standard
ISBD	<a href="#">International Standard Bibliographic Description (ISBD)</a>	Global	Structure Standard
LCNAF	<a href="#">Library of Congress Name Authority File</a>	Global	Data Value Standard
LCSH	<a href="#">Library of Congress Subject Headings</a>	Global	Data Value Standard
LOUD	<a href="#">Linked Art Data Model</a>	Global	Data Model
MARC	<a href="#">MACHINE Readable Cataloging</a>	Global	Structure Standard
METS	<a href="#">Metadata Encoding and Transmission Standard</a>	Global	Structure Standard

METS/ALTO	<a href="#">ALTO: Technical Metadata for Layout and Text Objects</a>	Global	Structure Standard
MODS	<a href="#">Metadata Object Description Schema</a>	Global	Structure Standard
PREMIS	<a href="#">Preservation Metadata Implementation Strategies</a>	Global	Data Dictionary
RAD	<a href="#">Rules for Archival Description</a>	National	Descriptive Standard
RDA	<a href="#">Resource Description and Access</a>	Global	Descriptive Standard
TGM	<a href="#">Thesaurus for Graphic Materials</a>	Global	Data Value Standard
VIAF	<a href="#">Virtual International Authority File</a>	Global	Data Value Standard
VRA Core	<a href="#">Visual Resources Association</a>	Global	Descriptive Standard

### 2.1.3 Data formats

‘The Cultural Heritage domain is characterised by many different types of data (audio-visual documents, written documents and those of intangible heritage – food, festivals, arts and crafts etc.), many knowledge needs and multiple communities of potential users.’<sup>16</sup> Yet when it comes to data formats for still images, what is produced and shared is remarkably homogeneous. The various guidelines produced for institutions engaging in digitisation typically recommend a limited number of formats for dissemination and preservation, and most institutions abide by these fairly strictly. Analogue materials (artefacts, objects, photographs, documents etc.) are commonly transformed through digitisation into static or ‘raster images’ represented by a fixed number of pixels. Born digital materials may be received as vector graphics ([EPS](#) or [SVG](#), for example) or in a wider variety of raster formats (camera RAW files unique to the camera manufacturer) and these might be made publicly available, but access images derived from these formats are always offered as well.

Access images shared in-browser are almost always [JPEG](#), while downloadable content may be either JPEG, [PNG](#) or [TIFF](#). High resolution images are envisioned as those needed for print and TIFFs are typically deemed to be the appropriate delivery format for publishers. A [PDF](#) might be produced in cases where text appears alongside the image, mainly to facilitate the use of Optical Character Recognition (OCR). All of these formats are widely used, well-documented open file formats that are actively monitored by the community for ongoing sustainability. The Library of Congress has published a [Recommended Formats Statement](#) that is supported by extensive research into the [Sustainability of Digital Formats](#). The National Archives in the UK undertakes similar work through [PRONOM](#), a technical registry of file formats, software products and other technologies related to the preservation of electronic records and digital content that is widely used for digital preservation software designed for the cultural heritage sector.

<sup>16</sup> European Commission, Directorate-General for Research and Innovation, Brunet, P., De Luca, L., Hyvönen, E., et al., (2022) *Report on a European collaborative cloud for cultural heritage : ex – ante impact assessment*, Publications Office of the European Union, <https://data.europa.eu/doi/10.2777/64014>.



Although not yet widely adopted and supported, there is strong advocacy for the creation of 3D images for cultural heritage, with 3D technologies being a particular focus of the *Commission Recommendation for a common European data space*. The EC advises that by 2030, Member States should digitise in 3D all monuments and sites identified as at risk and 50% of the most physically visited monuments, buildings and sites.<sup>17</sup> Similarly, in the United States the Smithsonian has begun actively working on a [3D Digitization](#), although the work is carefully positioned as that of a ‘small group of technologists.’ As of 2021, only 0.03% of Europeana’s assets were 3D files.<sup>18</sup>

Indications are that the purpose and value of 3D imaging in the image sharing landscape will probably be very different from traditional digital images. The Association for Library Collections and Technical Services advises in their *Minimum Digitization Capture Recommendations*:

Even with the rise of inexpensive 3D printers, one cannot currently make an adequate reproduction from a three-dimensional object, and so the intent is not to digitize so users have a surrogate to use, but rather to give the user general information about the object. Three-dimensional objects will most likely be reimaged at a later point.<sup>19</sup>

Somewhat in contrast to the position taken by the ALCTS, the [Task Force on 3D Content in Europeana](#) is more optimistic of the value of the 3D models being created for cultural heritage, although they also admit that this remains a developing field and decisions around 3D imaging are still largely dependent on both the tools for rendering and the purpose of the rendering for the end user. While the role of 3D images and the implications for their use is still being explored, the [Europeana 5Dculture project](#) is actively creating more 3D content and positioning the Europeana platform to encourage sharing and reuse.

The requirements for rendering 3D models in browsers and those for serving up 3D content for download vary, as do the range of tools in use by GLAMs for the creation of their 3D material. At this point it is unclear if there will be further implications for image sharing platforms beyond the technical rendering of the image files, as 3D images otherwise utilise the same metadata and discovery protocols as 2D raster image formats. Currently 3D images appear to be created primarily as reference materials for exhibition, teaching and learning rather than actual surrogates for cultural objects which seems to suggest there will be unique considerations for how 3D images are treated as both technologies and informational objects.

---

<sup>17</sup> 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>.

<sup>18</sup> European Commission (2021) *Commission proposes a common European data space for cultural heritage*.

<sup>19</sup> The Association for Library Collections and Technical Services Preservation and Reformatting Section (June 2013) *Minimum Digitization Capture Recommendations*,

<https://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>.

#### 2.1.4 Rights and licences

An analog object in a photo archive collection generally consists of far more than a single photograph. Rather, it is a compound and complex layered object, whether considered before or after its digitization. Rights may subsist or arise at any level, including the underlying work of art, the photograph of the work of art, the archival mount with its associated text and data, the digital reproduction of the photograph and/or the digital facsimile of the archival mount in full, and the data in the catalog record describing the object. In addition to copyright, which may apply to one or more of these layers, additional considerations include contractual restrictions, privacy implications, cultural permissions, and other ethical concerns.<sup>20</sup>

A single photograph may have a single copyright status, but a number of applicable licences, including but not limited to: IP, commercial, contractual, data protection and moral. The distinction between rights and licences is not always clearly articulated to the end user by the GLAMs and this ambiguity naturally carries over to the aggregator platforms, although each of the platforms reviewed for this document have made efforts to encourage institutions to apply licences to their content as well. Rights are usually clearly stated, while licences might be implied at best. Because of complications largely arising from acquisition processes that transfer the responsibility of stewardship, but not necessarily ownership of copyright to the collecting institution, the burden of responsibility is shifted to the end user to determine whether their desired reuse of an image is allowed.

This is not to say that usage licences are not suggested by the GLAMs at all, more so that they are unclear about how they communicate it and whether the licence is enforceable. It seems that the default usage licence implied by all institutions, regardless of the copyright status of the original object, is one that requires attribution and restricts commercial use. A courtesy statement is typically requested for any reuse of an image, and depending on the usage intent, a fee may be requested for high resolution files. This ensures that the use of any images can always be tracked back to the collecting institution, which can be an aid to discoverability, particularly for those institutions that do not have a robust digital collections platform or do not contribute to aggregator sites, and there are institutions that actively track the reuse of their collections. The fee for service model was also for many years positioned as a valuable revenue stream for GLAMs, though this perception seems to be largely falling out of favour as digitisation is fast becoming an integral part of collections work rather than a boutique, on-demand service offering.

---

<sup>20</sup> PHAROS Intellectual Property Working Group (2020) 'International Copyright Workshop,' [http://pharosartresearch.org/sites/default/files/pharos/files/pharos\\_international\\_copyright\\_workshop\\_report.pdf](http://pharosartresearch.org/sites/default/files/pharos/files/pharos_international_copyright_workshop_report.pdf).

The image platforms reviewed for this report are encouraging institutions to be much more transparent in their use of licences. Europeana requires a CCO licence for all metadata and advises all contributing institutions to apply [Creative Commons](#) licences to images wherever possible, however the use of [Rights Statements](#) by the Rights Statements Consortium is also advised and may be used alone in situations where CC licences cannot be legally applied. [Rightsstatements.org](#) is a global consortium, focusing specifically on the needs of the cultural heritage sector; members include DPLA, Europeana, National Digital Library of India and Trove, amongst others. The consortium has two working groups on technical implementation requirements and on usability and design principles.

Licences are not always assigned to metadata records, though interestingly this seems to be more of an oversight than a deliberate decision to restrict the reuse of the metadata. The Digital Public Library of America's guidance for data producers in its [Metadata Application Profile](#) states that the metadata provided represents objective fact and is therefore not copyrightable, implying that there is no licence required to reuse metadata from the cultural heritage sector.<sup>21</sup> This statement represents an interpretation of copyright legislation that will necessarily differ across legal jurisdictions. However, it may be likely that most GLAM institutions simply take for granted that their metadata is intended to do a public good and is therefore available to copy and reuse. The other image platforms reviewed have tried to avoid any broad assertions over whether metadata is legally in the public domain by relying on their Terms of Use to advise contributors of their options in making metadata available. DigitalNZ, in contrast to DPLA, advises in its [Shared Repository Terms of Use](#) that contributors may select both commercial and non-commercial licences for metadata.

## 2.2 Enabling technologies

The GLAM sector has for many years dealt with the interoperability of content as an analogue issue, which it has managed through the development of vocabularies and ontologies, descriptive standards and classification rules. As mentioned above, although the sector has historically prioritised the metadata as the material for exchange across digital systems, the need to facilitate access to digital image files as well as their metadata in machine-readable format has encouraged a wider uptake of exchange protocols and web technologies.

---

<sup>21</sup> In both Australia and Canada there have been some cases that indicate the contents of databases (directories, phone books etc.) are not copyrightable, but there does not appear to be any further evidence of this being applied to cultural heritage metadata. See 'Australian and New Zealand copyright law for databases, compilations and directories' (17 March 2011), AJPark, <https://www.ajpark.com/insights/australian-and-new-zealand-copyright-law-for-databases-compilations-and-directories/> and Emir Crowne, *Canadian Federal Court of Appeal rules that copyright does not subsist in real estate database*, *Journal of Intellectual Property Law & Practice*. Volume 13, Issue 6 (June 2018), 441–442, <https://doi.org/10.1093/jiplp/jpy046>.

### 2.2.1 Web technologies – HTTP, HTML, XML

Easy to overlook when it comes to enabling technologies are [HTTP](#) and [HTTPS](#), request-response protocols which enable communication on the web. Image sharing platforms largely make use of the latter, secure transfer protocol, to authenticate the accessed website and protect the integrity of the data exchanged. Both image files and associated metadata are presented in digital collections on the open web, primarily using [HTML](#), which structures the display of web content, and XML, which encodes information about resources in a machine-readable way.

### 2.2.2 Metadata harvesters – Z39.50, OAI-PMH

Many library catalogues expose bibliographic information via the [ANSI/NISO Z39.50](#) standard, a client/server protocol for searching and retrieving information from remote databases. The Z39.50 protocol has been exploited for both ingest and discovery, allowing libraries to copy catalogue records describing the same or similar resources while also pushing metadata to reference tools like Zotero or Refworks, and services such as interlibrary loan and online document delivery (examples include the Library of Congress Z39.50 [Gateway to Library Catalogs](#) and OCLC's [Z39.50 Cataloging service](#)).

The technological requirements for aggregating image metadata in the platforms used across the GLAMs have been largely facilitated by use of the OAI-PMH protocol (Figure 2). OAI-PMH allows repositories to expose their metadata to data harvesters and aggregator platforms as XML data. OAI-PMH data model requires the use of unqualified Dublin Core as a minimum, but otherwise allows content providers to create (and expose) additional metadata according to their own requirements and preferences. Given the heterogeneity of GLAMs, it is this flexibility which has led it to become the key enabling technology for metadata harvesting.

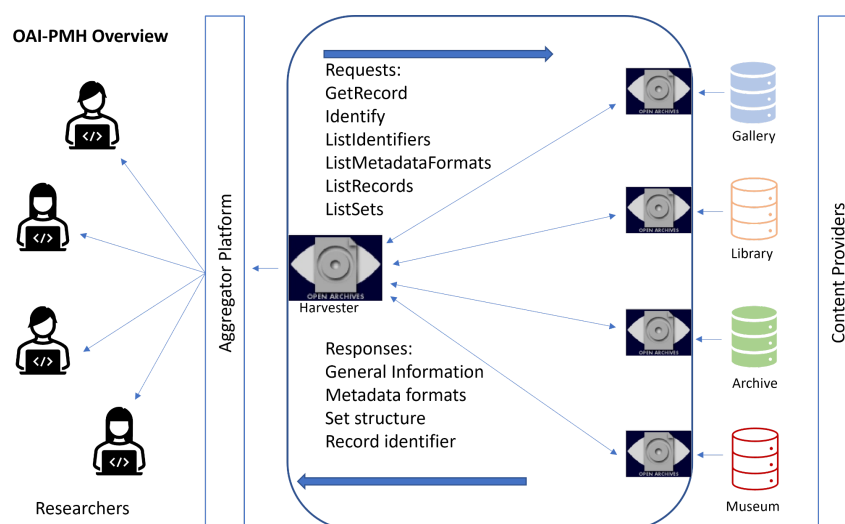


Figure 2 - OAI-PMH Overview. Source: [OAI-PMH](#), (Adapted by Joan Murphy), CC-BY-SA

Application Programming Interfaces (APIs) also play a significant part in the delivery of machine-readable metadata by allowing different computer programmes to communicate using a common language defining acceptable structures for the exchange of data. This method of communication allows questions to be asked and answers to be received in a way that is machine-readable and interpretable by humans. By enabling API usages, image sharing platforms can periodically harvest metadata with little human intervention and keep their content synchronised with that of the content provider. All of the platforms examined in this document go one step further by making their APIs open. Open APIs allow users to decide the parameters of queries using the protocol, with the DPLA describing their [Philosophy of Open APIs](#) as one which aims ‘to encourage the independent development of applications, tools, and resources that make use of data contained in the DPLA platform in new and innovative ways, from anywhere, at any time.’

### 2.2.3 Semantic web technologies

The semantic web as envisioned by Tim Berners-Lee, inventor of the World Wide Web, is a collection of structured data that can be linked and referenced by machines, creating a web of Linked Open Data. The mechanism by which this is implemented is RDF, a data modelling language which provides a simple framework to describe (and define) relationships between objects using the structure of triples (subject, predicate and object). RDF exists in a number of formats (RDF/XML, Turtle, JSON). The most widely used format in the GLAM sector is [JSON-LD](#), a linked data format based on JSON which is lightweight and human-readable and allows JSON properties to be mapped to ontological concepts. RDF requires the use of Uniform Resource Identifiers (URIs) for its successful implementation.<sup>22</sup>

The GLAM sector has a long history of creating identifiers to facilitate the retrieval of objects (including accession numbers, call numbers, barcodes etc.) and of information (all of the widely used metadata standards in the sector now provide URIs for elements and terms). For many years now, GLAM digital collections and online catalogues have offered ‘permalinks’ to resources, which are simply static URLs pointing to metadata records in digital collections. A persistent identifier (PID) is a special type of identifier, similar to a permalink, in that it not only provides a unique reference to objects and information on the web, but also ensures an unchanging and persistent reference is maintained (usually by a third party service). PID services offered by organisations such as DataCite and CrossRef have explicit schema requirements which distinguish them from URIs.<sup>23,24</sup> There is currently a handful of persistent identifier services supplying PIDs to the GLAM sector, though again the use of simple URIs to point to metadata records is much higher. Professionals in the archival

---

<sup>22</sup> RDF requires URIs while JSON-LD actually recommends [Internationalized Resource Identifiers](#) (IRIs), which is an expanded and modified protocol based on URI.

<sup>23</sup> DataCite Metadata Working Group (2021) ‘DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4’, p. 82 pages, <https://doi.org/10.14454/3W3Z-SA82>.

<sup>24</sup> ‘Schema versions - Crossref Documentation’ (5 October 2021) Crossref.org, <https://www.crossref.org/documentation/schema-library/schema-versions/>

sciences have developed their own distinct PIDs known as [Archival Resource Keys](#) (ARKs), which provide URLs that return specific metadata records through a persistent service.

Due to the complexity of the task, the level of connectivity and linking envisioned, the semantic web is still largely aspirational; however, significant efforts are being made by the GLAM sector to implement semantic web technologies with a growing number of GLAMs using services like Wikidata to expose and enrich their metadata. The [Wikidata](#) community in turn makes use of GLAM-related data value standards such as the [Virtual International Authority File \(VIAF\)](#) to link records. These efforts are supported by [W3C](#), the organisation that develops web standards and technologies to support a '[Web of data](#)'. It should be noted that all the metadata standards outlined in section 2.1.2.2. (Table 2) above are made available as LOD.<sup>25</sup>

#### 2.2.4 International Image Interoperability Framework (IIIF)

The [IIIF](#) provides a set of open standards widely supported by GLAMs. An international, community-driven initiative, it is leading the way in the implementation of open APIs for the exchange of cultural heritage images and associated metadata. At present it has 63 [partner organisations](#) including Europeana, the Getty Trust, major universities (Cambridge, Yale, Stanford) and the national libraries of Britain, France and the United States.

'IIIF is a way to standardize the delivery of images and audio/visual files from servers to different environments on the Web where they can then be viewed and interacted with in many ways.'<sup>26</sup> It offers [six open APIs](#) – the Presentation API, with its concept of a IIIF Manifest combining the individual elements (image, technical and descriptive metadata and URI) into a single object which enables an image viewer to understand how to process the data being made available to it. This standardised structuring of image data (as a JSON-LD file) implements LOD practices. By making these manifests publicly available, the images also become available through open web searches on sites such as Google.

For the researcher, IIIF implementation opens up a world of possibilities when it comes to viewing and comparing image content across repositories. The technology enables a standard view of high resolution image files, regardless of where the host server resides. The manifest structure allows image sharing platforms to display thumbnails/surrogates while also providing a URI that resolves to the content provider's image server. In this way, whether a researcher accesses an image via an aggregator or individual GLAM institution, they are guaranteed a high resolution file of verified integrity.

---

<sup>25</sup> 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>.

<sup>26</sup> 'How It Works,' IIIF, <https://iiif.io/get-started/how-iiif-works/>.

### 2.2.5 Embedded metadata

Image files often contain embedded metadata, which can range from basic technical information (e.g. information about how the file should be displayed or the camera settings used to capture the image) to descriptive bibliographic information about the image content.

The primary image formats containing embedded metadata are JPEG and TIFF, with both file encodings containing segments that allow for the addition of metadata tags. The most well-known type of embedded metadata is the [Exchangeable Image File Format](#) (EXIF) which includes information on camera and lens type, camera settings used to take the image (e.g. f-stop) and GPS coordinates. Other types of commonly embedded metadata include [IPTC](#) and [ICC](#) standards. In recent years the limitations of all these folder-based standards have been removed with the development of [eXtensible Metadata Platform](#) (XMP), which uses XML to extend the embedded metadata capabilities of JPEG and TIFF files and offers the greater flexibility for descriptive metadata (rights, bibliographic, licensing, keywords etc.). In addition, it can also be used in PNG, GIF and WebP files (as well as many other non-image file formats). XMP was first developed by Adobe in 2001 and, since 2012, Part One of the specification (the data model and namespaces) has been an open ISO standard, although other parts remain proprietary. The standard receives support from initiatives by Creative Commons, IPTC, PRISM and W3C (via RDF) and is based on the Dublin Core metadata element set.

There is a lot of overlap between the fields used by EXIF, IPTC and ICC, and many image files contain a combination of them all, however it is important to note that technical metadata is easily removed from files when they are copied into new file structures or transformed for online access. As some derivative files on image platforms are generated by the platform itself (particularly in the case of thumbnails for search and retrieval), they will often contain metadata that pertains to that particular file iteration/version and will not contain metadata associated with the source image file.

Embedded metadata seems to have uneven uptake across the cultural heritage sector, with little advice on the topic appearing in digitisation policy and guidance documents.<sup>27</sup> As part of their ingest process each content provider decides the workflow and technology that is best suited to their needs and therefore controls the types, and richness, of the metadata embedded.

## 2.3 Image sharing platforms

While the cultural heritage sector has long had systems in place for the exchange of information between catalogues, the need for centralised discovery platforms to improve exposure and access to unique collections has emerged alongside the increased demand for digitisation and digital

---

<sup>27</sup> For example, FADGI guidelines only advise embedding metadata in cases where an image of an archival object may be composed of multiple images stitched together, in which case the metadata describing the relationships of the parts to the whole must be recorded in the files themselves. Federal Agencies Digital Guidelines Initiative (FADGI) Still Image Working Group (January 2023) *Technical Guidelines for Digitizing Cultural Heritage Materials (Third Edition)*, <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html>.

collections. There are many platforms that have emerged to support the dissemination of cultural heritage image content, which can be organised generally into three broad categories:

1. Platforms that offer online exhibition-like functionality for the content provider ([Google Arts & Culture](#), [Omeka](#)),
2. Platforms that aggregate metadata from other repositories or catalogues ([Archives Hub](#), [Cultural Japan](#), [DigitalNZ](#), [DPLA](#), [Europeana](#), [Trove](#)),
3. Platforms that harvest content and also serve as repositories in their own right ([Wikimedia Commons](#), [Internet Archive](#), [Flickr](#)).

The first category described here offers the cultural heritage sector hosted services for disseminating digital collections and exhibitions, which often duplicate data and metadata available on other platforms in order to present the objects with a unique interpretive narrative. Google Arts & Culture offers a stylish interface for a content provider's data with a focus on exhibits which target a general public audience user experience. Omeka, on the other hand, is widely used by the research community as a content management system and publishing platform, but each Omeka contributor actually publishes a separate and standalone instance of the platform so that it is difficult to select one or two representative uses of the platform.<sup>28</sup> As such this category was seen to be outside the scope of our report and we have chosen to exclude it from our review.

Platforms that aggregate from other repositories allow images and their associated metadata to be discovered and accessed alongside the collections of other repositories. Many of these are national initiatives such as Archives Hub (UK), Cultural Japan (Japan), DigitalNZ (New Zealand), DPLA (United States) and Trove (Australia), with Europeana seeking to serve the wider European community.

The larger transnational platforms harvest content from multiple regions, and offer support for content providers ranging from individuals to national GLAM organisations. Crucially they also offer their own storage, serving as repositories independent of national initiatives, offering a 'consumer-to-consumer' type of service. Wikimedia Commons and the Internet Archive are the major platforms, making accessible all types of content (from web archives to books to films), alongside Flickr, where the focus is solely on images.

The advantage of image sharing platforms to researchers is that they offer a single search interface across multiple content providers, returning results using their own selected metadata fields. Most platforms will present high-level metadata on the results pages (title, basic description, content provider's name) requiring the researcher to click through to a content provider's site for more detailed information. Images themselves are returned as thumbnails, probably better described as access copies, and are almost always low quality JPEG renderings of the source image file (an exception to this is DigitalNZ which presents WebP format thumbnails).

---

<sup>28</sup> This is true for Omeka Classic and Omeka-S, which are downloadable open source software packages that can be hosted locally, as well as Omeka.net, a web hosting service that publishes Omeka collections as separate instances.



In general, image sharing platforms do not offer dynamic links to a content provider’s original metadata, which means that metadata updates must be pushed or manually updated on the platform for any changes to be reflected. While the IIF’s [Change Discovery API](#) does allow for dynamic updating of technical metadata (creation/deletion/move) it does not update descriptive metadata. Efforts are beginning to include synchronisation as part of harvesting workflows, with DPLA embarked on a project in April 2022 to continually [synchronise](#) its Wikimedia data.

In the following pages, the image delivery technologies and standards of six image sharing platforms are reviewed. The section concludes with a summary table of image sharing platforms (Table 3) for reference and comparison.

### 2.3.1 DigitalNZ

**Scope:** National      **Services:** Aggregator, Exhibition, Repository\*      **Images:** 4.7 million

#### Structure and governance

A [national governmental initiative](#), DigitalNZ was established in 2008 and is led by the National Library of New Zealand, reporting to the Dept of Internal Affairs. During its first decade it was guided by an advisory board, and is now an established part of the library with 6 core staff.

#### Development and sector support

The platform aggregates the content of 200+ organisations, at no cost to the content provider. It is widely supported and is an exemplar of a national digital strategy. \*Primarily an aggregator, the platform also encompasses a [repository](#) for at-risk digital content from small collections. It guarantees access to, but not preservation of, these materials.

#### Access

Access is free to all users, with the site offering registered users additional functionality such as the ability to upload content and access a personal API key.

#### Policies

The platform provides clear information on its policies such as [metadata](#), [copyright and licensing](#).

#### Rights/reuse

Clear information on rights and reuse for each image. CC licences supported. Platform metadata is freely available for all non-commercial reuse.

#### File formats

All image downloads are thumbnails in the WebP format (unique amongst platforms considered for this mapping report). Users are directed to the content provider site for access to image files, and clicking on the thumbnail brings the user directly to the source site. Users may upload content as to collections as JPEG, PNG and GIF (max file size 15MB)

#### Metadata - technical and descriptive

DigitalNZ has developed an open source tool called [Supplejack](#) to harvest data from its content providers, using HTML, RSS, XML, OAI-PMH and RDF/XML. This tool relies on the [Supplejack Metadata Schema](#) which uses the [DigitalNZ Metadata Dictionary](#) – Dublin Core is the underlying schema.

Their API is open and the platform also harvests data of relevance to New Zealand from sources with similarly open licences such as Europeana, Cultural Japan and DPLA. It supports LOD via RDF/XML.

### **Summary**

This platform offers an exemplar of a national image aggregator and demonstrates what is possible with a clear national strategy, with governmental support, clearly defined policies and investment in open source technology. Harvesting metadata from across New Zealand and presenting it in an accessible, user-friendly format, it has excellent functionality and follows best practice in web design. The development of the harvesting tool ensures wide collection, with clear information on licensing and rights in all cases.

### 2.3.2 DPLA

**Scope:** National

**Services:** Aggregator, Exhibition

**Images:** 12 million

#### Structure and governance

DPLA is a not-for-profit organisation funded by a number of foundations (incl.; Wikimedia, Mellon, McArthur) and membership fees. An independent organisation comprising multiple partners and hubs (often location-specific) across the library sector in the United States, it is governed by a board and employs a small number of staff. The DPLA membership governs the organisation's activities through advisory councils, working groups and task forces. Membership fees contribute toward the cost of metadata ingest and publication for individual organisations.

#### Development and sector support

Both a platform and a portal, the development of the DPLA began in 2010, with a vision of being an 'open, distributed national digital library to harness the power of collaboration to inform and empower everyone.'<sup>29</sup> The site itself went live in 2013 and is widely supported across the United States library landscape. In addition to metadata ingest it also offers peer support and development, and curates exhibitions from disparate collections on its platform. Membership is via Service Hubs (in effect sub-aggregators that provide content through a single feed from multiple partners) and Content Hubs with a direct relationship to DPLA. Both hub types provide metadata that resolve to a digital object.

#### Access

Content on the platform is universally accessible at no cost for viewing and download, membership is a requirement for uploading to the platform (via its 'hubs'). Open Access is facilitated by the provision of a freely available [API](#) to allow members access and reuse its enhanced metadata. DPLA offers a pipeline service to its members to allow them to contribute content to Wikimedia Commons (IIIF manifest required) and to date it has facilitated the [upload of 3m image files](#) to the Commons. It uses open web standards and best practice for accessibility, it currently has c12m images available to the end user.

#### Policies

Via the [DPLA Pro](#) Wiki, the platform makes available all of its [documentation](#) and tools with clear policies on metadata quality, licensing and copyright. It utilises both [rightsstatements.org](#) and Creative Commons and provides information on its collection development.

#### Rights/reuse

The platform only accepts content that is freely and openly available on the web. It also maintains a [Rights Portal](#) providing information to the wider US cultural heritage sector. As part of the ingest/upload process a URI to both a CC licence and a rights statement is required, and there is also a facility to provide additional rights metadata.

---

<sup>29</sup> DPLA (2023) *Strategic Plan, Digital Public Library of America*, <https://pro.dp.la/about-dpla-pro/strategic-plan>.

### **File formats**

The DPLA harvests and aggregates metadata, and does not, as such, have file format requirements. It requires a thumbnail to be delivered with each record, but the platform does not host content. Images can be downloaded from the site's curated collections as PNG and JPEG thumbnails but the user must go to the content provider site to access other formats. Thumbnails contain no embedded metadata (bar filename).

### **Metadata - technical and descriptive**

DPLA harvests descriptive metadata from content providers and generates additional information by way of an 'enrichment profile' (this often includes information such as geocoding, DC mappings etc.). Descriptive metadata, both that which is generated by DPLA, and that of its content providers, is designated as CC0 (Public Domain). The metadata required for harvesting from its hubs is very basic, with only 4 mandatory fields (Aggregation provider name, link to original record, resource rights, and resource title) although 7 others are '(strongly) recommended if available'. DPLA also permits harvesting, collection and modification of metadata from its own platform to others under CC0 licence.

Regardless of the content provider's underlying schema, all data is mapped to the [DPLA Metadata Application Profile](#) and stored and published using JSON-LD. As part of its work with Wikimedia, DPLA has altered its requirements, and data model, to allow for the inclusion of a URI. Where content is identified as being suitable for Wikimedia purposes a URI by way of a IIIF manifest is required. DPLA does not itself host a IIIF server.

DPLA both exposes and harvests its metadata using the [OAI-PMH](#) standard. In 2021 it began development of a model in Structured Data with Wikimedia Commons, this has led to DPLA being one of the first content providers to implement data synchronisation with the platform, making for a dynamic metadata record on Commons.

DPLA does not generate PIDs but the platform has a mandatory requirement for a URL linking to the content provider's instance of the image, which it displays as part of its own (i.e. DPLA's) metadata record.

### **Summary**

DPLA is a mature platform, with well-established and documented policies, procedures and workflows. Although situated in the library domain it offers examples of best practice that are applicable across the wider GLAM sector. Its work with Wikimedia puts it at the forefront of developments in the area of LOD, and its flexibility with regard to content provider metadata requirements makes it an excellent point of reference for the practical implementation of interoperability aspirations.

### 2.3.3 Europeana

**Scope:** International

**Services:** Aggregator, Exhibition

**Images:** 31 million

#### **Structure and governance**

An initiative of the European Union to support its cultural heritage sector in digital transformation, Europeana depends on a collaboration of three groups: the Europeana Foundation, a not-for-profit that operates the platform and infrastructure, the Europeana Network Association which is an open community of experts working in cultural heritage and the Aggregator's Forum, who represent the network of content providers. In addition to being an exhibition platform it operates as an advocacy group for open digital practices.

#### **Development and sector support**

The platform has wide sectoral support and is embedded in the digitisation strategies of many EU countries. Through EuropeanaPro it ensures buy-in by offering support with skills, training and resources for staff of content providers. Many countries have a single aggregator that feeds content directly from individual GLAM organisations. At its core Europeana has a small team that develops tools and services for the sector, with the aggregators ensuring that content is suitable for upload.

#### **Access**

Accessibility is one of the key tenets of the platform, and all content is freely available to the end-user. Content can be browsed with or without an account, account holders have additional benefits such as saving searches and creating their own public and private galleries. The platform offers the user the option to view the metadata in all official EU languages. Only registered Aggregators may upload to the platform.

#### **Policies**

The platform has clear policies with regard to [copyright and licensing](#) as well as its processes around [metadata and media resources](#).

#### **Rights/reuse**

Since 2012 Europeana's metadata has been available under CC0 licence, and any metadata submitted to it by content providers must adhere to the same CC0 licence. Images must select rights statements from either those available at Creative Commons or rightsstatements.org. This information is clearly displayed when an image is shown on the platform.

#### **File formats**

As an aggregator platform Europeana does not host image files and only thumbnails are required for ingest. The thumbnail (JPEG) can be downloaded from the platform, to access the original file the user must click through to the content provider's site.

#### **Metadata - technical and descriptive**

Europeana has its own mature and widely supported [data model](#) and content providers must submit metadata according to the specified format. Content providers' metadata is automatically

augmented by LOD (datasets, vocabularies) as part of the [Europeana Semantic Enrichment Framework](#). The platform displays key metadata and licensing information for each image with the user is also given the option of viewing more complete metadata on the platform.

Europeana is actively involved in the development of IIIF and is committed to its implementation, [extending its data model accordingly](#) by adding a IIIF profile. While not all content providers implement IIIF those that do provide a manifest as part of their ingest process and there are in excess of [300,000](#) IIIF records on the platform (including DRI).

Multiple [APIs](#) are freely available for download and the platform has clear information on accessing content SPARQL and LOD and data harvesting. Embedded EXIF metadata in thumbnails is minimal.

### Summary

Europeana underpins the common European Data Space for Cultural Heritage and is the go-to platform for the EU's digital cultural heritage. While it is by no means exhaustive it does provide access to all the key EU repositories via a single interface, which is of great benefit to the researcher. It has an active collaborative development base via [EuropeanaPro](#) and continues to drive sector best practice in Europe.

### 2.3.4 Wikimedia Commons

**Scope:** Global      **Services:** Harvesting, Repository, Dissemination      **Images:** [83 million](#)

#### Structure and governance

Wikimedia Commons is a free media repository, one of a number of projects and platforms hosted by the Wikimedia Foundation, a non-profit organisation founded in 2003.

Fundamentally a publishing platform, Wikimedia facilitates dissemination of images and associated metadata which are uploaded to its servers. It does not act as a portal to the content provider's own repository data or metadata. It does not expose the content provider's record, nor does it aggregate the metadata record from the content provider.

#### Development and sector support

With 500+ staff and a global network of volunteer contributors the platform has emerged as a robust and reliable source of copyright-free content. In addition to accepting content from individual users the platform has a number of partnerships in place with key image repositories such as the Library of Congress and the Wellcome Trust, and many GLAM institutions contribute content on a project and/or *ad hoc* basis such as the Smithsonian, British Library and Rijksmuseum.

#### Access

Content on the platform is universally accessible at no cost for viewing and download, however user registration is a requirement for uploading to the platform. Open Access is facilitated by the provision of multiple freely available [APIs](#).

#### Policies

It has developed and published clear [policies and guidelines](#) around copyright, licensing, deletion and formats, and also provides specific support to smaller GLAM organisations via its [GLAM wiki project](#).

#### Rights/reuse

The platform only accepts reusable content, either that which is in the Public Domain or under Creative Commons 0, BY or BY-SA licences. When downloading content an unambiguous attribution statement is provided to the end-user. The licence requirement explicitly applies to both metadata and image file.

#### File formats

The platform only accepts non-proprietary image file formats, stating a [preference](#) for JPEG, PNG and SVG, although GIF, TIFF, WebP and XCF are also present. The majority of downloads/exports are as JPEG with the platform generating multiple resolution options, from thumbnail to original image, for the end user to choose from. Although not a preferred format it should be noted that the platform also hosts a large number of TIFF files (primarily from cultural heritage institutions like the Library of Congress) which are downloadable by the user.



### Metadata - technical and descriptive

Using APIs the content provider populates prescribed Wikimedia data fields. Wikimedia has its own [data model](#) with JSON and RDF serialisations (i.e. Linked Open Data support). While the ontology is defined in a table of [properties](#) it does not provide a mapping to any standards, such as Dublin Core or FOAF, at present.

Embedded EXIF metadata is [supported](#) both at point of import and export and this data can be accessed by an end user with appropriate EXIF viewing software. Depending on the source organisation's practices this can include information such as rights, institutional IDs and bibliographic catalogue information.

At present there is no official roadmap for default integration with IIF, however recent enhancement to the Wikimedia APIs allows for technical developments and interconnections with the IIF manifest. The IIF community also operates a Slack Wikimedia channel and there are [synergies](#) between the two communities. Wikimedia does operate a IIF server, although this is not widely promoted.

There is no requirement for a PID to be associated with the data, and where a PID does exist there is no requirement for it to be exposed as part of the Wikimedia metadata record.

### Summary

Wikimedia Commons is currently the most accessible repository for image files, allowing end-users access to multiple resolutions and embedding possibilities for reusable images from a broad range of GLAM sector content providers. It adheres to many Open Science principles with clear policies, guidelines and standards, and an active user community. Development of IIF integration, associated PIDs, and projects working to dynamically update metadata via APIs (e.g. an ongoing project with [DPLA](#)) can only serve to enhance its status as a research resource.

### 2.3.5 Internet Archive

**Scope:** Global      **Services:** Harvesting, Repository, Dissemination      **Images:** [4.5 million](#)

#### **Structure and governance**

A not-for-profit organisation based in the US, the Internet Archive began in 1996, primarily to archive websites. It has since expanded its remit to include digital versions of other types of content, including images. It is funded by a number of foundations (Mellon, NSF, Knight) and is also a member of the [DLF](#), [DPC](#) and [ALA](#), among many other grassroots cultural heritage groups. Another funding stream is provided by their web archiving service (Archive-It).

#### **Development and sector support**

The platform offers a number of services to the GLAM sector (including [Archive-It](#), which allows organisations to archive their own websites) with content hosted and stored on the IA's servers. Currently circa 240 content providers use the service as an exhibition/dissemination platform for its archival collections (including [NASA](#)).

#### **Access**

All content on the platform is free to access for the researcher, users also have the option to register for a free account, allowing them to upload content to the platform.

#### **Policies**

The IA does not clearly state its policies on its website beyond the [Terms of Use](#).

#### **Rights/reuse**

While the implicit message is that all content that is available for download is in the public domain there is no explicit statement to this effect. Image files are accepted only if the content provider confirms that they have the right to share the content. The upload process includes the option to add licence metadata (although the field is not mandatory) and supports CC licences.

#### **File formats**

Support is provided for GIF, JPEG, PNG, TIFF and JP2 image files. Images can be downloaded in multiple formats.

#### **Metadata - technical and descriptive**

According to its website the IA is '[metadata agnostic](#)' and as such it has a wide-ranging [schema](#). It assigns its own unique identifiers to each image file. The IA has a [Developer Portal](#) which makes available all APIs. Embedded EXIF data if present is available to the end user.

At present there doesn't appear to be any support or development for LOD and while IA implements IIIF, this is currently at the beta stage of [development](#).

### Summary

The Internet Archive is an accessible platform for researchers, however it does not impose criteria that validate the accuracy of its metadata. This lack of standards puts the burden on the researcher to ensure the quality of the images and metadata being engaged with. While some institutional content providers take great care with the material provided there is no consistency across the platform.

### 2.3.6 Flickr

**Scope:** Global      **Services:** Harvesting, Repository, Exhibition platform      **Images:** 5 billion

#### **Structure and governance**

Primarily known as a for-profit organisation with an engaged user-community, Flickr.com is membership-based with two offerings, free accounts that allow for limited content upload, and unlimited download and Pro membership which entails a fee but offers unlimited storage, file back-up and metrics software. Flickr Commons began as a project of Flickr with the stated goal of cataloguing the world's public photo archives, with approximately 125 participating institutions (including Smithsonian, [Library of Congress](#), National Library of Ireland, UK National Archives), all of which have Pro membership. The responsibility for the support and development of Flickr Commons has now been transferred to the not-for-profit Flickr Foundation (with a landing site at [Flickr.org](#)) - a new initiative with an Executive Director and small advisory board (currently recruiting its first staff members (Jan 2023)) looking to develop a long-term outlook for Flickr Commons content.

#### **Development and sector support**

Many major GLAM organisations use Flickr as a exhibition platform but it also provides smaller GLAM organisations with a cost effective dissemination and back-up/off-site storage facility. Flickr actively engaged with the GLAM sector to develop the Commons and is currently implementing its Flickr Commons Revitalization Plan: 2021-2023 by establishing the Flickr Foundation. Very much in the early stages, it has an ambitious strategy to develop a [100-year Challenge](#) to secure the future of Commons content. The high-profile academic partners in this enterprise include Melissa Terras, Jane Winters, Tedi Odumosu and Eliza Gregory.

The sections below outline what is *currently* available to the research community via Flickr.com, it is likely that the development of Flickr Commons by Flickr.org will introduce enhanced structures and standards with more rigorous requirements.

#### **Access**

An open API that is free to use for members and developers/integrators for non-commercial purposes.

#### **Policies**

Clear policies with regard to licensing and copyright, no guidance for metadata.

#### **Rights/reuse**

Content ranges from All Rights Reserved to Public Domain, but the platform encourages use of CC licences. All images contain explicit licensing terms.

### **File formats**

The platform accepts JPEG, PNG, GIF (non-animated), up to a maximum file size of 200MB for ingest. Members may set different parameters for downloads, and many GLAM content providers allow for end-users to export multiple resolutions in JPEG format.

### **Metadata - technical and descriptive**

Content providers can elect to show or hide EXIF data at ingest, and may then make further decisions about whether to allow it to be retrieved by end-users. Where privacy settings allow, embedded metadata is visible to end-users. This XMP data can include EXIF, IPTC and descriptive metadata such as bibliographic data.

There is no information available on LOD, IIIF etc.

### **Summary**

The development of Flickr Commons, by way of the Flickr Foundation, promises much for the future of this image sharing platform. It seems likely that Flickr Commons will exist as a platform similar in scope to Wikimedia Commons, providing robust schema and interoperability mechanisms, as well as adopting principles such as the [CARE Principles for Indigenous Data Governance](#). For this reason Flickr should continue to be viewed as an important player in the cultural heritage image sharing landscape. As-is, the current Flickr.com platform may not provide a high degree of quality metadata; however, as a research tool, the ability to build on the Flickr API allows technically aware users to create powerful searches across a vast dataset (5 billion images).

**Table 3 - Image Sharing Platforms - Summary Table**

	DigitalNZ	DPLA	Europeana	Wikimedia Commons	Internet Archive	Flickr
<b>Geo Scope</b>	National	National	European	Global	Global	Global
<b>Repository</b>	Yes*	No	No	Yes	Yes	Yes
<b>Exhibition <sup>a</sup></b>	Yes	Yes	Yes	No	No	Yes
<b>Aggregator <sup>b</sup></b>	Yes	Yes	Yes	No	No	No
<b>No. of Images</b>	4.7m	12m	31m	83m	4.5m	5 billion
<b>Open APIs</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Viewing version</b>	Access/ Thumbnail	Access/ Thumbnail	Access/ Thumbnail	Full Res	Full Res	Full Res
<b>Export Format</b>	WebP	JFIF	JPEG	JPEG**	JPEG**	JPEG**
<b>EXIF/XMP</b>	No	No	No	Yes	Yes	Yes
<b>Data Model</b>	Supplejack	DPLA-MAP	EDM	Wikibase	Unclear	Unclear
<b>Metadata Standard</b>	DC	DC	DC	DC	DC	Unclear
<b>LOD support <sup>c</sup></b>	Yes	Yes	Yes	Yes	No	No
<b>IIIF <sup>d</sup></b>	Unclear	Yes	Yes	Yes	Yes	No
<b>Creative Commons <sup>e</sup></b>	Yes	Yes	Yes	Yes	Yes	Yes

<sup>a</sup>Exhibition

The platform provides additional services such as curated exhibitions

<sup>b</sup>Aggregator

Platform that links back to content providers' site for full image and metadata

<sup>c</sup>LOD support

Linked Open Data supported/on roadmap, but not necessarily implemented

<sup>d</sup>IIIF support

IIIF is supported but not necessarily implemented for all content providers

<sup>e</sup> 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

### 3. Summary of key issues raised for the WP13 case study

This report provides an overview of the ways in which image sharing platforms that support the cultural heritage sector have leveraged existing practice in the field to reach a broad community of end users. It further provides the foundation for connecting this work with the specific needs of researchers operating both within aligned academic domains of research (mainly the Humanities and Social Sciences) and more broadly across the public and private research sectors via the FAIR principles, which will be explored in the next deliverable. Below is a high-level summary of DRI's practices related to those reviewed in section 2 of this report, followed by conclusions and next steps.

#### 3.1 Positioning DRI in the landscape

The Digital Repository of Ireland's own 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 the image sharing platforms reviewed in this report.

*Acquisition and Selection for Digitisation:* The DRI operates as a membership-based organisation and generally deposits are made by DRI members or those affiliated with DRI member organisations. A fee is charged for membership, although this fee is waived for applicants seeking to deposit through the [Community Archive Scheme](#). In addition to the membership requirement, DRI also has a Collections Development Policy that provides an overview of the collecting areas and types of data considered appropriate for the repository.<sup>30</sup> The DRI also publishes and preserves collections and data assembled in the course of its own collaborative research projects (e.g. the Wellcome Trust funded [Archiving Reproductive Health](#) research project is responsible for several digital collections in the repository). The repository supports either partial or full removal of metadata and data from the repository in its Withdraw Data Policy.<sup>31</sup> As both a repository and a research-driven collecting institution, DRI's organisational model is probably more aligned with the GLAM organisations that contribute to the image sharing platforms than the platforms themselves.

*Discovery:* Collections and metadata ingested into the repository are made freely accessible, with restrictions on usage licences applied only where depositing institutions deem necessary. The repository supports the export of data in the [BagIt](#) format, and provides an OAI-PMH feed for metadata harvesting as either native Dublin Core or conforming to the Europeana Data Model. An [API](#) is also available for developers wishing to interact with metadata in the repository. DOIs are issued for every collection and every object record through [DataCite](#). Collections information and images may also be published online by the depositing institutions, though links are not provided to original records from the DRI. Typically, depositors with the DRI use the repository for both access and preservation, and do not duplicate collections on their own platforms. The DRI is an accredited [national aggregator](#) for the Europeana platform.

<sup>30</sup> Digital Repository of Ireland (2021) *DRI Collection Policy*, <https://doi.org/10.7486/DRI.kk91v774c-2>.

<sup>31</sup> Digital Repository of Ireland (2018) *DRI Withdraw Data Policy*, <https://doi.org/10.7486/DRI.0r96mp375>.

*Description:* The DRI's repository architecture currently only allows for two levels of hierarchy in collection arrangement: collection and object. The collection record largely facilitates transparency around the context for objects entering the repository. In that respect, the DRI's conceptual model uses the collection record effectively as a provenance record, analogous to a dataset's README file or supporting documentation. Provenance information is also minimally available within the object record, usually in the fields allocated to Rights and Depositing Organisation. Data and metadata are functionally represented in a 1-to-1 relationship (one image gets one record) with no further dependencies.

The use of the Dublin Core metadata standard as the default record entry allows for considerable flexibility in mapping to common descriptive metadata standards, which is likely why DC is also the preferred standard for the image sharing platforms reviewed. DRI supports ingest of metadata in numerous formats including Dublin Core, Qualified Dublin Core, EAD, MODS or MARC XML formats and this metadata is made available to download as XML. Contributors may select from a set of recommended data value standards when ingesting, or they may suggest an addition to the DRI's current list. The DRI does not currently support LOD and there is no external linking from subject headings or other controlled terms in the metadata record.

*Level of Openness:* The DRI does not currently assign a licence to the metadata records in the repository, although provisions in the Depositor Terms and Conditions do reserve the right for the repository to make available certain metadata fields [Title, Creator, Publisher, Publication Year and DOI] under a public domain dedication.<sup>32</sup> Licences for digital objects in the repository are otherwise assigned to the data by the depositors, and content may be restricted where deemed appropriate. Content on the DRI's main website is otherwise made available under a CC BY licence. This lack of transparency around licences does not extend to rights, which are always clearly articulated with a dedicated rights statement. Citations are available formatted according to three different bibliographic standards (MLA, APA and Chicago), but the default suggestion does not follow any particular formal citation style. This option instead aligns with common cultural heritage practices in prioritising the collection name, stewarding institution(s) and relevant identifiers in the following format: Collection. (Date) Object title, Repository [Distributor], Depositor [Depositing Institution], DOI.

*Digitisation Technical Standards:* The DRI supports the ingest, display and preservation of the common file formats for still images utilised by the cultural heritage sector, and displays these images natively in browser alongside the metadata record or in a IIIF viewer. Users may download the access images and, if allowed by the depositor, the original image file. The repository does not currently embed descriptive metadata into access images created by the repository, although image files provided by institutions may contain this information if it is included in the files deposited for publication. There is currently limited support for some 3D file formats, but work is planned to offer

---

<sup>32</sup> Digital Repository of Ireland (2018) *DRI Deposit Terms and Conditions*, <https://doi.org/10.7486/DRI.1544r4085>



more accessible visualisations for these file types in the near future. 3D files are described and presented with the same metadata record format as two-dimensional images.

*Preservation:* The DRI is a CoreTrustSeal-certified repository which ensures the active management of digital content for long-term preservation and access (see the DRI's statement on [Digital Preservation](#)). DRI produces and retains metadata about technical preservation actions carried out on image files but does not make this information publicly available. The DRI does allow depositors to edit and update metadata records associated with their collections and does not maintain a public record of version changes.

### 3.2 Conclusions and next steps

The DRI's support for image data sharing in the cultural heritage sector will be reviewed by WP13's Working Group on Cultural Heritage Image Sharing, which includes a range of professionals in the field with considerable experience in diverse areas of policy, technology and practice. The next deliverable will also review and expand on a growing body of literature on FAIR alignment in the field (for instance, Europeana's [Guidelines to FAIRifying Data](#) and the 'Always Already Computational: Collections as Data Final Report'<sup>33</sup>).

Of course, cultural heritage has only recently begun to think of collections *as data* at all, such that the theme of an upcoming conference for the Digital Research Infrastructure for the Arts and Humanities (DARIAH) this year will be '[Cultural Heritage Data as Humanities Research Data?](#)' The presence of the question mark in that statement points to very real concerns over how a shift to the language of data may create or reinforce old power dynamics and undermine emerging efforts to provide sensitive, community-centred description. Unpacking the implications of versioning metadata changes will be an important area of work in the recommendations report.

Additional important takeaways from this report include:

- ❖ Data formats used by the GLAMs are open, limited in number and widely accessible. 3D images are an emerging data format that may require different policies, as well as technologies, to make them accessible and preservable. Characterising them appropriately as data, or data structures that deliver data and metadata, may be a challenge.
- ❖ Metadata interoperability is well-developed and supported by a variety of technologies, standards, crosswalks and data models. There is, however, a tendency towards reducing both the granularity of the metadata and complexity of the metadata structures in order to facilitate this, which can mean that there is richer metadata available on organisational websites than on the image sharing platforms.
- ❖ Data interoperability is facilitated by a limited number of Web technologies, with IIIF perhaps the most significant new development in the field.

---

<sup>33</sup> Thomas Padilla et al. (2019) 'Always Already Computational: Collections as Data: Final Report' *Copyright, Fair Use, Scholarly Communication, etc.*. 181. <https://digitalcommons.unl.edu/scholcom/181>

- ❖ URIs (expressed as URLs) are favoured over DOIs and other PIDs, and have been successfully used to support both persistent data retrieval and the use of LOD.
- ❖ While copyright is generally stated clearly, data and metadata usage licences are sometimes unclear or more restrictive than the copyright status might suggest. It is generally implied that data usage without attribution is not allowed, although image sharing platforms encourage open licences for metadata.
- ❖ Provenance information relating to the stewardship of collections is generally available, but limited. Information about the acquisition and ongoing care of digital collections (both original objects and their surrogates) is not usually made available.
- ❖ Despite established and robust practices and policies for digital preservation in the sector, there is no requirement for organisations or image sharing platforms to maintain either data or metadata records as originally published, and there is a noticeable lack of user-facing transparency around administrative or preservation actions.

## 4. Bibliography

American Library Association, MARBI Committee (1996) 'The MARC 21 Formats: Background and Principles'. Library of Congress. Available at: <https://www.loc.gov/marc/96principl.html> (Accessed: 2 February 2023).

Association for Library Collections and Technical Services Preservation and Reformatting Section (June 2013) *Minimum Digitization Capture Recommendations*. Available at: <https://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>. (Accessed: 17 February 2023).

AJPark (17 March 2011) 'Australian and New Zealand copyright law for databases, compilations and directories.' Available at: <https://www.ajpark.com/insights/australian-and-new-zealand-copyright-law-for-databases-compilations-and-directories/> (Accessed: 17 February 2023).

Berthiaume, G. 'If You Want to Go Far, Go Together: The Collaboration among the GLAM Community in Canada (2016–2019).' *Research Library Issues*, no.300 (2020): 6–17. Available at: <https://publications.arl.org/rli300/1>. (Accessed: 23 January 2023).

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. Available at: <https://doi.org/10.3233/SW-210425>.

Crowne, E. (June 2018) 'Canadian Federal Court of Appeal rules that copyright does not subsist in real estate database,' *Journal of Intellectual Property Law & Practice*. Volume 13, Issue 6 (441–442). Available at: <https://doi.org/10.1093/jiplp/jpy046>.

DataCite Metadata Working Group (2021) 'DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4', p. 82 pages. Available at: <https://doi.org/10.14454/3W3Z-SA82>.

Digital Preservation Coalition (2015) *Digital Preservation Handbook*. Available at: <https://www.dpconline.org/handbook/contents> (Accessed: 2 February 2023).

Digital Repository of Ireland (2021) *DRI Collection Policy*. Available at: <https://doi.org/10.7486/DRI.kk91v774c-2>.

Digital Repository of Ireland (2018) *DRI Deposit Terms and Conditions*. Available at: <https://doi.org/10.7486/DRI.1544r4085>.

Digital Repository of Ireland (2018) *DRI Withdraw Data Policy*. Available at: <https://doi.org/10.7486/DRI.Or96mp375>.

DPLA (2023) *Strategic Plan, Digital Public Library of America*. Available at: <https://pro.dp.la/about-dpla-pro/strategic-plan>. (Accessed: 23 January 2023).



DPLA (n.d.) *Philosophy, Digital Public Library of America*. Available at: <https://pro.dp.la/developers/philosophy> (Accessed: 20 February 2023).

Erway, R. and Schaffner, J. (2007) 'Shifting Gears: Gearing Up to Get Into the Flow'. OCLC. <https://www.oclc.org/content/dam/research/publications/library/2007/2007-02.pdf>. (Accessed: 23 January 2023).

European Commission (2019) 'DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC,' *Official Journal of the European Union*, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>

European Commission (2021) *Commission proposes a common European data space for cultural heritage | Shaping Europe's digital future*. Available at: <https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage> (Accessed: 2 February 2023).

European Commission, Directorate-General for Research and Innovation, Brunet, P., De Luca, L., Hyvönen, E., et al., (2022) *Report on a European collaborative cloud for cultural heritage : ex – ante impact assessment*, Publications Office of the European Union, <https://data.europa.eu/doi/10.2777/64014>.

Europeana (2021) *Semantic Enrichment Framework, Google Docs*. Available at: [https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit?usp=embed\\_facebook](https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit?usp=embed_facebook) (Accessed: 25 January 2023).

Federal Agencies Digital Guidelines Initiative (FADGI) Still Image Working Group (January 2023) *Technical Guidelines for Digitizing Cultural Heritage Materials (Third Edition)*. <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html> (Accessed: 3 February 2023).

Gregory, A. and Hodson, S. (2022) 'WorldFAIR Project (D2.1) "FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?"' Available at: <https://doi.org/10.5281/ZENODO.7378109>.

Harpring, P. (revised 25 January 2022) 'Metadata Standards Crosswalk,' Getty Research Institute, [https://www.getty.edu/research/publications/electronic\\_publications/intrometadata/crosswalks.html](https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html).

'How It Works,' IIF, <https://iif.io/get-started/how-iif-works/> (Accessed: 17 February 2023).

Jyrkkiö, T., ed. (2010) *Long-term Loans Best Practice Report*. Finnish National Gallery. [https://www.ne-mo.org/fileadmin/Dateien/public/topics/Collection\\_Mobility/Long-term\\_loans\\_best\\_practices\\_report.pdf](https://www.ne-mo.org/fileadmin/Dateien/public/topics/Collection_Mobility/Long-term_loans_best_practices_report.pdf). (Accessed: 23 January 2023).

OCLC (2018) *Best practices for creating sharable metadata, OCLC Support*. Available at: [https://help.oclc.org/Metadata\\_Services/CONTENTdm/Get\\_started/best\\_practices](https://help.oclc.org/Metadata_Services/CONTENTdm/Get_started/best_practices) (Accessed: 3 February 2023).

Padilla, T. et al. (2019) 'Always Already Computational: Collections as Data: Final Report' *Copyright, Fair Use, Scholarly Communication, etc.*. 181. <https://digitalcommons.unl.edu/scholcom/181>



PHAROS Intellectual Property Working Group (2020) 'International Copyright Workshop'. Available at: [http://pharosartresearch.org/sites/default/files/pharos/files/pharos\\_international\\_copyright\\_workshop\\_report.pdf](http://pharosartresearch.org/sites/default/files/pharos/files/pharos_international_copyright_workshop_report.pdf) (Accessed: 20 January 2023).

Proffitt, M. (2008) 'Impact of Digitization on Scholarship and Collecting,' *Hanging Together*, the OCLC Research Blog. Available at: <https://hangingtogether.org/impact-of-digitization-on-scholarship-and-collecting/> (Accessed: 25 January 2023).

'Schema versions - Crossref Documentation' (5 October 2021) Crossref.org, <https://www.crossref.org/documentation/schema-library/schema-versions/> (Accessed: 17 February 2023).

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

Zierau, E. (2013) 'PREMIS, METS and preservation metadata': Available at: <https://www.dpconline.org/docs/miscellaneous/events/856-premismets-zierau-1-1/file>. (Accessed: 23 January 2023).

Zorich, D., Waibel, G. and Erway, R. (2008) *Beyond the Silos of the LAMs: Collaboration Among Libraries, Archives and Museums*. OCLC Research. Available at: <https://doi.org/10.25333/X187-3W53>.

#### 4.1 List of web links (in alphabetical order)

<http://ala.org/>

[https://api.wikimedia.org/wiki/API\\_catalog](https://api.wikimedia.org/wiki/API_catalog)

<https://archive-it.org/>

<https://archive.org/about/terms.php>

<https://archive.org/details/image>

<https://archive.org/details/nasa>

<https://archive.org/developers/metadata-schema/index.html#metadata-schema>

<https://archive.org/developers/tutorial-get-ia-credentials.html>

[https://archivescanada.ca/wp-content/uploads/2022/08/RADComplete\\_July2008.pdf](https://archivescanada.ca/wp-content/uploads/2022/08/RADComplete_July2008.pdf)

[https://archivesforblacklives.files.wordpress.com/2020/11/ardr\\_202010.pdf](https://archivesforblacklives.files.wordpress.com/2020/11/ardr_202010.pdf)

<https://archiveshub.jisc.ac.uk/>

<https://artsandculture.google.com/>

<https://blog.flickr.net/en/2021/06/21/cataloging-covid-19-with-the-library-of-congress/>

<https://cidoc-crm.org/crmdig/sites/default/files/CRMdigv4.0.pdf>

[https://commons.wikimedia.org/wiki/Commons:Exif#Purpose\\_for\\_using\\_Exif\\_at\\_Commons](https://commons.wikimedia.org/wiki/Commons:Exif#Purpose_for_using_Exif_at_Commons)

[https://commons.wikimedia.org/wiki/Commons:International\\_Image\\_Interoperability\\_Framework](https://commons.wikimedia.org/wiki/Commons:International_Image_Interoperability_Framework)  
[https://commons.wikimedia.org/wiki/Commons:Policies\\_and\\_guidelines](https://commons.wikimedia.org/wiki/Commons:Policies_and_guidelines)  
[https://commons.wikimedia.org/wiki/Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)  
[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Properties\\_table](https://commons.wikimedia.org/wiki/Commons:Structured_data/Properties_table)  
[https://commons.wikimedia.org/wiki/Main\\_Page](https://commons.wikimedia.org/wiki/Main_Page)  
<https://commons.wikimedia.org/wiki/Special:MediaStatistics>  
<https://creativecommons.org>  
<https://cultural.jp/>  
<https://datacite.org/>  
<https://digitalnz.github.io/supplejack/api/creating-schemas.html>  
<https://digitalnz.org/>  
<https://digitalnz.org/about/our-team>  
<https://digitalnz.org/about/terms-of-use/shared-repository-terms-of-use>  
<https://digitalnz.org/developers/supplejack>  
<https://digitalnz.org/make-it-digital/describing-digital-content/metadata-resources>  
<https://digitalnz.org/make-it-digital/enabling-use-re-use>  
<https://digitalnz.org/partners/shared-repository>  
<https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit>  
[https://docs.google.com/document/u/0/d/1Z3I\\_ckQWjnQQ4SzpORbClcIXUheO-Jd4jt-oZFuMcoQ/pub](https://docs.google.com/document/u/0/d/1Z3I_ckQWjnQQ4SzpORbClcIXUheO-Jd4jt-oZFuMcoQ/pub)  
<https://doi.org/10.5281/zenodo.4608430>  
<https://dp.la/>  
<http://dp.la/info/developers/map/>  
<https://dp.la/news/wikimedia-project-update>  
<https://dpo.si.edu/3d-digitization-program>  
<https://dri.ie/dri-community-archive-scheme>  
<https://dri.ie/digital-preservation>  
<https://dri.ie/europeana>  
[https://drive.google.com/file/d/1fJEWnYy5Ch7\\_ef\\_-V48-FAViA72OieG/view?pli=1](https://drive.google.com/file/d/1fJEWnYy5Ch7_ef_-V48-FAViA72OieG/view?pli=1)  
[https://en.wikipedia.org/wiki/Wikipedia:Image\\_use\\_policy#Format](https://en.wikipedia.org/wiki/Wikipedia:Image_use_policy#Format)  
<https://en.wikipedia.org/wiki/HTTPS>  
<https://en.wikipedia.org/wiki/HTTP>

[http://files.archivists.org/groups/museum/standards/8.%20Workflow%20Documentation/BHS%20Digitization\\_Cataloging\\_Manual\\_BrooklynHistoricalSociety.pdf](http://files.archivists.org/groups/museum/standards/8.%20Workflow%20Documentation/BHS%20Digitization_Cataloging_Manual_BrooklynHistoricalSociety.pdf)

<https://guides.dri.ie/developer-guide/01-api-doc.html>

<https://icom.museum/en/research/cultural-democracy-and-inclusion/>

<https://icom.museum/en/resources/standards-guidelines/objectid/>

<https://icom.museum/en/research/sustainability-and-local-development/>

[https://icom.museum/wp-content/uploads/2018/09/CP\\_Statement-independence-of-museums\\_EN.pdf](https://icom.museum/wp-content/uploads/2018/09/CP_Statement-independence-of-museums_EN.pdf)

[https://icom.museum/wp-content/uploads/2022/02/Accessioning-Standards\\_EN.pdf](https://icom.museum/wp-content/uploads/2022/02/Accessioning-Standards_EN.pdf)

[https://icom.museum/wp-content/uploads/2022/02/Key\\_concepts\\_EN.pdf](https://icom.museum/wp-content/uploads/2022/02/Key_concepts_EN.pdf)

<https://id.loc.gov/authorities/names.html>

<https://iiif.archivelab.org/iiif/documentation#overview>

<https://iiif.io/>

<https://iiif.io/api/discovery/1.0/>

<https://iiif.io/api/index.html>

<https://iptc.org>

<https://jpeg.org/about.html>

<https://json-ld.org/spec/latest/json-ld-api-best-practices/>

<https://linked.art/model/>

<https://memory.loc.gov/ammem/about/techStandards.pdf>

<https://ndsa.org/publications/levels-of-digital-preservation/>

<http://omeka.net>

<http://openarchives.org/pmh/>

[https://outreach.wikimedia.org/wiki/GLAM/Get\\_started](https://outreach.wikimedia.org/wiki/GLAM/Get_started)

<https://pro.dp.la/about-dpla-pro>

<https://pro.dp.la/developers/api-codex>

<https://pro.dp.la/developers/philosophy>

<https://pro.dp.la/hubs/documentation>

<https://pro.dp.la/hubs/metadata-application-profile>

<https://pro.europeana.eu/>

[https://pro.europeana.eu/files/Europeana\\_Professional/Event\\_documentation/Webinars/ECC\\_Webinar\\_FAIRyfyngData.pdf](https://pro.europeana.eu/files/Europeana_Professional/Event_documentation/Webinars/ECC_Webinar_FAIRyfyngData.pdf)

<https://pro.europeana.eu/page/apis>

<https://pro.europeana.eu/page/copyright-and-digital-cultural-heritage>

<https://pro.europeana.eu/page/edm-documentation>  
<https://pro.europeana.eu/post/iiif-adoption-by-europeana-future-perspectives-for-the-network-1>  
<https://pro.europeana.eu/post/iiif-datasets-in-europeana-a-scholar-s-delight>  
<https://pro.europeana.eu/project/3d-content-in-europeana>  
<https://pro.europeana.eu/project/5dculture-deploying-and-demonstrating-a-3d-cultural-heritage-space>  
<https://pro.europeana.eu/share-your-data/process>  
<https://project-thor.readme.io/docs/introduction-to-persistent-identifiers>  
<https://repository.ifla.org/bitstream/123456789/697/1/digitization-projects-guidelines.pdf>  
<https://repository.ifla.org/handle/123456789/1183>  
<https://repository.ifla.org/handle/123456789/1842>  
<https://repository.ifla.org/handle/123456789/1955>  
<https://repository.ifla.org/handle/123456789/2026>  
<https://repository.ifla.org/handle/123456789/2146>  
<https://rights-portal.dp.la/>  
<https://rightsstatements.org/en/about.html>  
<https://saa-ts-dacs.github.io/>  
<https://trove.nla.gov.au/>  
<https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html>  
<https://viaf.org>  
<https://vraweb.org/resources/cataloging-cultural-objects/>  
<https://www.adobe.com/acrobat/about-adobe-pdf.html>  
<https://www.adobe.com/ie/creativecloud/file-types/image/raster/tiff-file.html>  
<https://www.adobe.com/ie/creativecloud/file-types/image/vector/eps-file.html>  
<https://www.adobe.com/ie/creativecloud/file-types/image/vector/svg-file.html>  
<https://www.adobe.com/products/xmp.html>  
<https://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations>  
<http://www.archivists.org/prof-education/pre-readings/IMPLP/AA68.2.MeissnerGreene.pdf>  
<https://www.cidoc-crm.org/>  
<https://www.color.org/index.xalter>  
<https://www.dariah.eu/2022/10/26/save-the-date-for-the-dariah-annual-event-2023/>  
<http://www.dlib.org/dlib/november07/kaufman/11kaufman.html>  
<https://www.diglib.org/>



<https://www.dpconline.org/>  
<https://www.dri.ie/archiving-reproductive-health#administrative-collections>  
<https://www.dublincore.org/>  
<https://www.flickr.com/>  
<https://www.flickr.org/about-us/the-100-year-plan/>  
<https://www.flickr.org/announcing-the-flickr-foundation/>  
[https://www.getty.edu/research/publications/electronic\\_publications/cdwa/index.html](https://www.getty.edu/research/publications/electronic_publications/cdwa/index.html)  
<https://www.getty.edu/research/tools/vocabularies/aat/>  
<https://www.getty.edu/research/tools/vocabularies/cona/index.html>  
<https://www.getty.edu/research/tools/vocabularies/ulan/>  
<https://www.gida-global.org/care>  
<https://www.ica.org/en/archival-arrangement-and-description>  
<https://www.ica.org/en/digital-imaging-photographic-collections-foundations-technical-standards>  
<https://www.ica.org/en/digital-records-curation-programme>  
<https://www.ica.org/en/guidelines-selection-digital-heritage-long-term-preservation>  
<https://www.ica.org/en/ica-copyright-declaration>  
<https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>  
<https://www.ica.org/en/preservation-issues-digitizing-historical-photographs>  
<https://www.ica.org/en/preserving-archives>  
<https://www.ica.org/en/principles-access-archives>  
<https://www.ica.org/en/reference-dossier-archival-claims>  
<https://www.ica.org/en/technical-guidance-managing-archives-restrictions-0>  
<https://www.ietf.org/rfc/rfc8493.txt>  
<https://www.json.org/json-en.html>  
<https://www.loc.gov/aba/cataloging/subject/>  
<https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>  
<https://www.loc.gov/aba/rda/>  
<https://www.loc.gov/cds/downloads/FRBR.PDF>  
<https://www.loc.gov/ead/>  
<https://www.loc.gov/marc/>  
<https://www.loc.gov/marc/bibliographic/>  
<https://www.loc.gov/marc/dccross.html>

<https://www.loc.gov/preservation/digital/formats/>  
<https://www.loc.gov/preservation/digital/formats/fdd/fdd000146.shtml>  
<https://www.loc.gov/preservation/resources/rfs/stillingm.html>  
<https://www.loc.gov/rr/print/tgm1/>  
<https://www.loc.gov/standards/alto/>  
<https://www.loc.gov/standards/mets/>  
<https://www.loc.gov/standards/mods/>  
<https://www.loc.gov/standards/premis/v3/>  
<https://www.loc.gov/standards/vracore/>  
<https://www.mediawiki.org/wiki/Wikibase/DataModel#Notes>  
[https://www.museums.ch/fr/assets/files/dossiers\\_d/Publikationen/NEMO\\_Encouraging\\_Collections\\_Mobility\\_A4.pdf](https://www.museums.ch/fr/assets/files/dossiers_d/Publikationen/NEMO_Encouraging_Collections_Mobility_A4.pdf)  
<http://www.nationalarchives.gov.uk/pronom/>  
<https://www.niso.org/publications/ansiniso-z3950-2003-s2014>  
<https://www.oclc.org/content/dam/research/activities/rights/practice.pdf>  
<https://www.oclc.org/research/activities/photoscan.html>  
<https://www.oclc.org/research/activities/pmwg/background.html>  
<https://www.oclc.org/research/activities/rights.html>  
<https://www.oclc.org/research/areas/data-science/fast.html>  
<https://www.oclc.org/research/areas/data-science/iiif.html>  
<https://www.oclc.org/research/publications/2021/oclcresearch-retrospective-accessioning-preprint.html>  
<https://www.oclc.org/research/publications/2022/strategic-collaboration/strategic-collaboration-report.html>  
<https://www.oclc.org/research/publications/books/library-linked-data-in-the-cloud/chapter2.html#x1-170002>  
<https://www.oclc.org/research/publications/library/visguides/visguide1.html>  
<https://www.openarchives.org/OAI/openarchivesprotocol.html>  
<https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html>  
<https://www.w3schools.com/html/>  
[https://www.w3schools.com/whatis/whatis\\_http.asp](https://www.w3schools.com/whatis/whatis_http.asp)  
[https://www.w3schools.com/xml/xml\\_what.asp](https://www.w3schools.com/xml/xml_what.asp)  
[https://www.w3.org/egov/wiki/Linked\\_Open\\_Data](https://www.w3.org/egov/wiki/Linked_Open_Data)  
<https://www.w3.org/International/O-URL-and-ident.html>  
<https://www.w3.org/RDF/>

<https://www.w3.org/standards/semanticweb/>

<https://www.w3.org/TR/png/>

<https://www.w3.org/TR/rdf-concepts/#section-Concepts>

<https://www.worldcat.org>

<https://www2.archivists.org/groups/intellectual-property-working-group/orphan-works-statement-of-best-practices>

<https://www2.archivists.org/groups/intellectual-property-working-group/well-intentioned-practice-for-putting-digitized-collections-of-unpublished-materials->

<https://www2.archivists.org/groups/standards-committee/open-archives-initiative-protocol-for-metadata-harvesting-oai-pmh>

<https://www2.archivists.org/groups/standards-committee/open-archives-initiative-object-re-use-and-exchange-oai-ore>

[https://www2.archivists.org/sites/all/files/GuidelinesForReappraisalDeaccessioning\\_2017.pdf](https://www2.archivists.org/sites/all/files/GuidelinesForReappraisalDeaccessioning_2017.pdf)

[https://www2.archivists.org/sites/all/files/BrooklynHistoricalSociety\\_DigitalPreservationPolicy\\_0.pdf](https://www2.archivists.org/sites/all/files/BrooklynHistoricalSociety_DigitalPreservationPolicy_0.pdf)

<https://www2.archivists.org/standards/ACRL-RBMS-SAA-Guidelines-on-Access-to-Research-Materials-in-Archives-and-Special-Collections-Libraries-2020>

<https://www2.nau.edu/libnap-p/protocols.html>