



## UvA-DARE (Digital Academic Repository)

### FAIR Principles for Library, Archive and Museum Collections

*A proposal for standards for reusable collections*

Koster, L.; Woutersen-Windhower, S.

**Publication date**

2018

**Document Version**

Final published version

**Published in**

Code4Lib Journal

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Koster, L., & Woutersen-Windhower, S. (2018). FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections. *Code4Lib Journal*, 40. <http://journal.code4lib.org/articles/13427>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections

*Many heritage institutions would like their collections to be open and reusable but fail to achieve that situation because of organizational, legal and technological barriers. A set of guidelines and best practices is proposed to facilitate the process of making heritage collections reusable. These guidelines are based on the FAIR Principles for scholarly output (FAIR data principles [2014]), taking into account a number of other recent initiatives for making data findable, accessible, interoperable and reusable. The resulting FAIR Principles for Heritage Library, Archive and Museum Collections focus on three levels: objects, metadata and metadata records. Clarifications and examples of these proposed principles are presented, as well as recommendations for the assessment of current situations and implementations of the principles.*

by Lukas Koster, Saskia Woutersen-Windhauer

### Current Issue

[Issue 41, 2018-08-09](#)

### Previous Issues

[Issue 40, 2018-05-04](#)

[Issue 39, 2018-02-05](#)

[Issue 38, 2017-10-18](#)

[Issue 37, 2017-07-18](#)

[Older Issues](#)

### For Authors

[Call for Submissions](#)

[Article Guidelines](#)

## Introduction

Most libraries, archives, museums (LAM) and other heritage institutions manage a variety of collections. These collections can be physical and digital book and journal collections, repositories of publications, physical artifacts collections and digital representations of these, archives of all kinds of physical and digital material, etc. The collections and objects are described in a large variety of information systems and databases, using all kinds of digital object and metadata formats, access protocols and storage facilities.

More and more people and organizations want to be able to access and reuse both digital objects and the (meta)data describing the physical and digital objects, not only individually but also in bulk. Some LAM institutions already comply with these wishes. Examples are: the British Library Free Data Services (British Library Free Data Services [2018]), the Ghent University Library Open Data services (Universiteitsbibliotheek Gent Open Data [updated 2018]) and the Rijksmuseum API (Rijksmuseum API [2018]). In other cases, there may be technical or legal barriers or data services lack clear documentation making discovery and access ambiguous. Not all collections apply or provide the universal metadata standards and protocols that are necessary for exchanging or reusing information in a simple way. Furthermore, some collections lack essential metadata (like global persistent identifiers), making implementation of universal standards and protocols impossible. And reuse can be complicated, as data or data services may lack information related to data reuse or license restrictions. Not only are these circumstances an obstacle for reuse of collection objects and metadata by external parties, but also for internal and external interoperability between various information systems used within LAM institutions. To summarize: it is not enough to make collections available through web based end user interfaces and provide download options for individual objects and metadata records. On the contrary: collection data and objects must be findable, accessible, interoperable and reusable for people and software in their entirety and in specific parts.[1]

In this article[2] we propose a body of requirements for making LAM collections findable, accessible, interoperable, and reusable (FAIR), as well as a set of recommendations for assessing existing institutional situations and for implementation of the requirements. These requirements and recommendations are based on the FAIR Principles for scholarly output (FAIR data principles [2014]) and adapted for cultural heritage collections.

FAIR collections are not necessarily Open collections. The Open Knowledge Foundation applies a definition of “openness” that adds modification permission and universal participation to the accessibility and reusability criteria, “*at no more than a reasonable reproduction cost*” (What is open? [2018]), referring to the Open Definition. The short form of the Open Definition is “*Open data and content can be freely used, modified, and shared by anyone for any purpose*”. The full definition stresses openness in license (public domain or open license), in access (on the web and free of charge) and formats (without restrictions on its use and processable with at least one free/libre/open-source software tool) (Open Definition 2.1 2018). Accessibility and reusability in the FAIR sense do not require collections and objects to be freely available, modifiable and shareable with free tools as such. That is why we have chosen to use the term ‘FAIR collections’ instead of ‘Open collections’. Some metadata or objects will be copyright protected, have privacy issues or local law issues. Local law issues are for instance Database rights in the European Union[3] or portrait right as a personality right in the Netherlands[4]. It can be discussed if open can be used for the metadata and FAIR is only necessary for the objects. Unfortunately, there are still vendors that protect data that is (or should be) in the public domain by a contract. For that reason, metadata cannot always be open access or be freely reused. The same applies to objects that should be, according to the copyright law, in the public domain. Even in those cases an owner can, as the owner of the object or metadata, lay down the conditions and determine what a user may or may not do with the data or objects.

## Current initiatives

---

In recent years there have been various local and global collaborative initiatives to develop sets of guidelines and best practices for making objects and associated metadata open and reusable. Many of these initiatives focus on open and reusable research datasets. But there are also initiatives that relate to cultural heritage objects and collections. A number of these initiatives are described below. First, we will look at the FAIR Data principles, then focus on a number of other initiatives in comparison to FAIR and finally we will discuss the respective relevance of all initiatives for reusability of LAM collections.

## FAIR

In 2014 the FAIR Data Principles were proposed as “*a concise and measurable set of principles*” to support “*the reuse of scholarly data*” (Wilkinson et al. 2016):

### **To be Findable:**

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

### **To be Accessible:**

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
  - A1.1 the protocol is open, free, and universally implementable.
  - A1.2 the protocol allows for an authentication and

*authorization procedure, where necessary.  
A2 metadata are accessible, even when the data are no longer available.*

**To be Interoperable:**

*I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.  
I2. (meta)data use vocabularies that follow FAIR principles.  
I3. (meta)data include qualified references to other (meta)data.*

**To be Re-usable:**

*R1. meta(data) have a plurality of accurate and relevant attributes.  
R1.1. (meta)data are released with a clear and accessible data usage license.  
R1.2. (meta)data are associated with their provenance.  
R1.3. (meta)data meet domain-relevant community standards.*

These principles describe a set of guidelines to make data (meaning ‘scholarly output’, more specifically ‘research datasets’) and metadata about these datasets Findable, Accessible, Interoperable and Reusable (FAIR). The authors stated that “... *the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals*” (Wilkinson et al. 2016). The FAIR principles do not require long term sustainable storage and access of datasets, but only long term sustainable access of metadata: “*A2 metadata are accessible, even when the data are no longer available.*”.

The phrasing of the FAIR principles is somewhat confusing, very probably because of the wish to be concise. Some principles (not all) refer to both “data” and “metadata”, which is formulated as “(meta)data”, for example in “F1. (meta)data are assigned a globally unique and persistent identifier”. One principle is self-referencing (“*I2. (meta)data use vocabularies that follow FAIR principles*”). Also the enumeration of the four sections (F, A, I, R), each containing 3 or 4 principles, is rather uncommon, for example “(F)air” has four individual principles F1, F2, F3, F4, but “(R)eusable” has one main principle R1 and three sub-principles R1.1, R1.2, R1.3. The logic of this subdivision is unclear, because all four Reusable principles are guidelines on the same level in their own right. In fact, the authors say so themselves: “The elements of the FAIR Principles are related, but independent and separable.” (Wilkinson et al. 2016, p.4).

## FAIR Annotations

The FAIR principles can be applied to other object types. Collocated with the FORCE2017 conference in Berlin, a meeting was organized by the Annotating All Knowledge Coalition dedicated to FAIR Annotations (Martone 2017). It is an initiative to make scholarly Annotations on the web findable, accessible, interoperable and reusable. A list of issues (under construction) is available (at the time of writing) in the form of discussions of all FAIR principles applied to annotations (Annotations and FAIR [2018]). The list raises several questions, for instance how to link annotations to specific fragments of a text, what a retrieval protocol for annotations should look like, etc.

## PARTHENOS

PARTHENOS is a consortium of national and international research and research infrastructure organizations, such as INRIA, KNAW, CLARIN, DARIAH. “*PARTHENOS aims at strengthening the cohesion of research in the broad sector of Linguistic Studies, Humanities, Cultural Heritage, History, Archaeology and related fields through a thematic cluster of European Research Infrastructures, integrating initiatives, e-infrastructure and other world-class infrastructures, and building bridges between different, although tightly, interrelated fields.*” (PARTHENOS/About the

project/The goal [2018]). PARTHENOS aims to “*provide common solutions to the definition and implementation of joint policies and solutions for the humanities and linguistic data lifecycle*”. For this end PARTHENOS “*will deliver guidelines, standards, methods, services and tools to be used by its partners and by all the research community*”.

In their “*Report on Guidelines for Common Policies Implementation (1)*” (PARTHENOS 2017), PARTHENOS presented a set of high-level principles as common guidelines to their community. These guidelines are based on the FAIR Principles, but also extend these. For instance, FAIR does not directly address long term preservation of scholarly output as such, but only long-term preservation of the metadata of those objects. PARTHENOS explicitly formulates policies, strategies and guidelines for long term digital preservation of objects (PARTHENOS 2017, par. 3.3.2). It refers to the much used OAIS (Open Archival Information System) standard (ISO [updated 2012]).

## Metadata2020

Metadata2020 (metadata2020.org) is a collaboration of stakeholders who create or use scholarly metadata. These stakeholders are service providers of platforms and tools, publishers, librarians, researchers, data publishers, data repositories and funders. With Metadata2020 they advocate richer, connected, and reusable, open metadata for research output with the aim to improve the quality of metadata for research. They make users and creators aware that metadata should be richer to fuel discoverability and innovation, to bridge the gaps between systems and communities, and reusable to be able to eliminate duplication of effort. Richer metadata of the Metadata2020 initiative fits all aspects of the FAIR principles ‘Findable’ and ‘Reusable’ (especially R1). Improving the connection between systems and communities is related to FAIR Principle ‘Interoperable’ (I1, I2 and I3) and ‘Reusable’ (R1.3). The FAIR principle ‘Reusability’ is also one of the three pillars of Metadata2020 (most relevant are the FAIR principles R1.1. and R1.2).

Some special challenges between researchers and libraries are mentioned by Metadata2020, such as that libraries have invested in siloed metadata standards, seem hesitant to develop new systems that support new models and prefer to fit new models into legacy systems (Metadata2020 Librarians [2018]). This has led to interoperability problems, and duplication of data entry for the use with different systems. Another challenge is that librarians use metadata in a different way compared to researchers, and that even different research fields use different metadata within their own systems. Libraries can play an important role if they work with the other stakeholders to build a common metadata vocabulary, including the nuances per research field. Libraries can also help with the adoption of persistent identifiers to prevent disambiguation of names and institutions. To conclude, libraries, together with the other stakeholders, should work towards open, de-siloed metadata that promotes cross-platform interoperability and sharing for all stakeholders.

## Cultural Heritage Reuse Charter

PARTHENOS, together with a number of other European institutions (CLARIN and DARIAH, both partners in PARTHENOS; APEF, Europeana, E-RIHS) participates in the initiative to set up a ‘Cultural Heritage Reuse Charter’. A draft Mission Statement was published in 2017 (APEF et al. [2018]) and a survey was held to collect feedback on the Mission Statement (CLARIAH [updated 2017]). The Mission statement formulates six generic principles that both heritage institutions and researchers should subscribe to: *Reciprocity, Interoperability, Citability, Openness, Stewardship, Trustworthiness*.

The subject of these principles is a bit unclear. Throughout the statement the phrase “*cultural heritage data*” is used without explaining the meaning of “*data*”. Sometimes “*content and knowledge*” is used, “*digital and non-digital assets*”, “*research data*”, “*output*”, “*data, digital collections or data descriptors*”.

Of these six principles *Reciprocity* seems somewhat out of scope. It states that “*Both Cultural Heritage Institutions and Researchers agree to share content and knowledge equally with each other, making use of data centres and research infrastructures*” and as such it merely serves as a recommendation of signing a mutual contract. In fact, all principles are two-sided and describe obligations for institutions as providers and researchers as consumers of heritage data. *Interoperability* is a combination of part of the FAIR principles ‘Interoperability’ and ‘Accessibility’: “*international standards, frameworks and interoperability protocols*”.

The *Citation* principle states that heritage institutions should apply human and machine readable “*relevant data citation standards*”. It can be questioned that this really something institutions should provide. It seems that citation standards can be derived from provided standard metadata formats as facilitated by tools like Refworks, Mendeley and Zotero.

The *Openness* principle only addresses the use of licenses, which is represented by the FAIR *Reusable 1.1* principle. Openness usually also comprises Findability and Accessibility principles, such as the use of open standards and protocols, persistent identifiers and online availability, for instance as formulated in Tim Berners-Lee *Five Stars of Open Data* (Berners-Lee [updated 2015]). The *Stewardship* principle focuses on long term preservation and accessibility. *Trustworthiness* refers to Provenance (FAIR principle *Reusable 1.1*), including descriptions of “*any relevant materials, equipment, techniques, procedures and protocols used*”. The final Trustworthiness requirement “*alert to any defamatory use of Cultural Heritage data*” seems rather out of scope and hardly feasible. It does not describe guidelines for presentation of content and data, but for monitoring usage of these.

## DEN – DE BASIS

DEN, the Dutch knowledge centre for digital heritage and culture, has published a set of minimal requirements for digitization of cultural heritage, in Dutch “DE BASIS” (DEN [updated 2016]).<sup>[5]</sup> It consists of six areas of attention: *Rights Management, Findability, Creation, Presentation, Digital Sustainability/Preservation, Description*. DE BASIS is a mixture of guidelines, principles, policies, references and roadmaps. As such it serves in fact as a set of principles combined with detailed roadmaps how to implement these principles.

- *Rights Management* extensively covers FAIR 1. (*meta*)data are released with a clear and accessible data usage license.
- *Findability* actually focuses on four issues: *Identification* (data are unique and persistently identified: FAIR F1, F4, A1), *Accessible* (data are accessible via the Internet: FAIR A1), *Search Machine Searchable* (more or less FAIR F, A, plus HTML search machine recommendations), *Reuse* (FAIR A1, A1.1, I1, I3, explicitly not the same as FAIR Reusability).
- *Presentation* covers the separation of content and presentation, look and feel, multilinguality, web accessibility
- *Digital Sustainability/Preservation* focuses on the long term preservation and accessibility of digital collections (FAIR does not really cover this as stated before)
- *Description* covers all types of metadata about digital objects (FAIR F1, F2, F4,

I1, I3, R1, R1.3).

## Discussion

In their brief existence, the FAIR Data Principles have experienced widespread publicity, adoption and adaptation in the scholarly information world and beyond (Mons et al. 2017). Some of the initiatives discussed here explicitly mention the FAIR principles and their adoption and adaptation of them. In other, older initiatives FAIR principles can be partly identified “after the fact”. Whereas FAIR and most of its emulators are targeting the scholarly output environment, some initiatives relate to heritage collections, which is the focus of this article.

Can the FAIR principles be transposed and applied to heritage collection objects? Not as such, but they are a good starting point. The two main limitations are the lack of explicit attention for long term preservation of digital objects, besides their metadata, and the excessive interwovenness of “data” (or objects) and “metadata”.

The FAIR-like initiatives focusing on heritage collections differ substantially, which is caused by differences in target audience and objectives. The Cultural Heritage Reuse Charter is in fact a mutual contract between institutions and researchers, and for that reason cannot be implemented unilaterally by institutions. The DEN DE BASIS set of best practices, although it is very broad and complete, is not compact enough for quick adoption.

We deem it necessary for collections to be FAIR. When collections do not meet the FAIR criteria, these collections will be harder to use by teachers, researchers and enterprises. Collections that will not be FAIR will run the risk of being out of the picture, as will in the long term their institution. The objective of this article is to provide a compact, and practical list of guidelines for achieving reusability of LAM collections that can be fairly easily applied by LAM institutions according to their own roadmaps and resources.

## Minimum requirements for objects and metadata

---

The FAIR Principles for scholarly output can be applied to LAM collection objects with a small number of adjustments and extensions, including some principles from the other initiatives mentioned here. The FAIR principles for LAM Collections are applied to three levels instead of two:

1. Objects (such as books, journals, artifacts, videos, datasets, etc.)
2. Metadata about the objects on elementary level (such as title, creator, identifier, date, etc.)
3. Metadata records (body of metadata elements about an object in a specific database)

## FAIR principles for objects

- Findable
  - Objects (physical and digital) have a globally unique persistent identifier
  - Objects are described with metadata
- Accessible

- Digital objects are permanently accessible by:
  - Sustainable storage (hardware, storage medium)
  - Open universal access protocols
  - Version management
  - Backups
  
- Interoperable
  - Digital objects are stored in preferred or acceptable formats
  
- Reusable
  - Digital objects have a date-timestamp
  - Objects have a license for reuse, which is also available in a machine readable form

## **FAIR principles for metadata**

- Findable
  - Metadata specify the global persistent identifier (PID) of the object. This PID is used in all systems/databases that contain a description of the object in question
  - Metadata about specific objects are available via one or more searchable online repositories, catalogs, online databases, etcetera
  
- Accessible
  - Metadata specify information about an object's availability, obtainability and/or access options
  
- Interoperable
  - Metadata are available at least in one metadata schema appropriate for the specific type of object
  - Metadata are available in various additional generic standard data formats for other contexts
  - Metadata contain links/references to other objects/authority files, by using other global persistent identifiers
  
- Reusable
  - Metadata specify the object's rights holder
  - Metadata contain license information referring to the object
  - Metadata specify the object's provenance



## FAIR principles for metadata records

- Findable
  - Metadata records have their own global persistent identifier
  
- Accessible
  - Metadata records are machine readable
  - Metadata records are accessible using open universal protocols
  
- Interoperable
  - Metadata records must be of sufficient quality
  
- Reusable
  - Metadata specify the metadata record's provenance
  - Metadata specify the entity responsible for the metadata record
  - Metadata records have their own licence for reuse, which is also available in a machine-readable form

## **Clarification of the FAIR principles for LAM collections**

---

### Persistent identifiers

By “persistent identifier” (PID) we mean a “globally unique persistent identifier”. An “identifier” is a code which can be used to refer to an object. “Unique” implies that the identifier is used exclusively for one particular object. “Global” means that the identifier is valid throughout the entire world wide web. “Persistent” entails the permanent availability of the identifier, independently of any individual organizations, systems or systems implementations.

Several systems are available for assigning global persistent identifiers to objects, such as Handle (handle.net), DOI (doi.org), URN-NBN (Hakala and Network Working Group [updated 2001]). For references to people (as author, subject etc.) a number of identifier systems can be applied, such as ORCID (orcid.org), ISNI (isni.org), VIAF (viaf.org). For indexing terms (subject headings, thesaurus terms, classification codes etc.) many identifier systems are available too, such as Library of Congress Subject Headings (Library of Congress Subject Headings [2018]), AAT (Getty Research Institute [updated 2017]), MeSH (National Center for Biotechnology Information, [2018]), etc. Besides persistent identifiers for objects it is also recommended to assign identifiers to metadata records describing the objects. The metadata records in question must contain the identifier of the described objects as one of its metadata elements. Both types of persistent identifier should be findable and usable via the online user interfaces of all systems and repositories that the metadata records are published in. It is still widespread common practice to use internal dedicated system identifiers which refer to both the metadata record and the object described. This type of internal identifier can be published on the web using a URL based on the particular system or domain, but it is not a global and persistent identifier, because it depends on the specific system or domain.

## Sustainable storage

It is vital that digital objects are stored in stable server environments in order to assure their long-term availability. Servers have to be managed, monitored and secured in a professional manner. Whenever necessary the objects have to be migrated to new environments. A professional backup system must be in place to avoid the loss of objects. A versioning system must be applied in order to maintain previous versions of newer replacements. These versions must be assigned date-timestamps in order to distinguish between different versions. Objects must be stored in preferred or acceptable formats (Recommended formats statement [2016]), such as PDF (text), JPEG, PNG (image), WAVE, FLAC (audio).

## Access protocols

For access to both digital objects and digital metadata, universal access protocols must be made available on several levels. The first level comprises server level access protocols for copying files between servers and workstations (FTP/SFTP), direct server command line access (SSH) and downloading files and data through web browsers (HTTP). The second level consists of application level access protocols for accessing metadata, such as application programming interfaces (APIs), harvesting (OAI-PMH) and retrieving metadata (Z39.50, SRU) and linked data (SPARQL). Finally, there are the general formats and frameworks, other than the domain specific data formats (see *Standard metadata formats* below), such as XML, RDF, JSON, JSON-LD, which are used to represent the domain specific formats. For instance, MARC can be represented as MARC-XML, Dublin Core in XML, RDF, etc. In order to support reuse of metadata and digital objects as much as possible, options for selecting and downloading in bulk must be presented as well as options for retrieving and downloading records and objects individually.

## Standard metadata formats

It must be possible to provide metadata records in various standard formats, other than the format the metadata is stored in. Examples are MARC, Dublin Core, MODS, METS. The stored metadata must be sufficient to describe the specific object type and support the intended reuse. In the linked data context vocabularies or ontologies in RDF are used. For this various standards are available, such as EDM (Europeana Data Model), CIDOC-CRM (for cultural heritage), BIBFRAME (the prospective linked data successor to MARC), as well as well-known formats like Dublin Core and schema.org.

## Copyrights and licenses

In order to avoid the recurring necessity of requesting permission for reuse in every individual case, it should be instantly clear what actions are permitted with the objects and metadata. This is possible by including a license provided by the rights holder. The most used open content license is the Creative Commons (CC) license.[6] The CC license is a worldwide, irrevocable, non-exclusive license for the duration of copyright and similar rights, and sui generis database rights. The most used version is the 'CC-BY' (Creative Commons. Attribution 4.0 International [2018]). With a CC-BY, the rights holder gives permission that metadata or an object may be distributed, copied and adapted free of royalties under the condition of attribution. In addition there are 3 other elements that can be added: ND (the metadata or object may not be adapted); SA (must be distributed under the same license); and NC (may not be used for commercial purposes). If copyright no longer applies, or if the copyright is waived by the rights holder, a CC-0 statement can be used to explicitly state that the object or metadata is a part of the public domain or that all copyright and related rights to the

fullest extent allowed by law are waived and can be used by anyone without restrictions (CC0 [2018]).

If a license cannot be granted for any reason, consideration can also be given to providing a standardized statement, such as 'Unknown rightsholder'. These standards can be found on the site [rightsstatements.org](https://rightsstatements.org) that has been set up by Europeana and the Digital Public Library of America (DPLA). Organizations will have to determine which license for reuse will be granted to which objects and metadata. In addition, organizations must take into account other rights such as contract, portrait, personality and privacy rights in relation to objects and metadata.

## External links

For the description of entities related to the objects (people/creators, subjects, organizations, places etc.) the accepted commonly used persistent identifiers, available in accepted external authority files, must be recorded in the metadata whenever possible. For people for instance ORCID, ISNI, VIAF etc. For subjects many sources are available, such as Library of Congress Subject Headings, Getty Art & architecture Thesaurus AAT, Medical Subject Headings MESH, etc.

## Provenance

Provenance originally stands for "*the chronology of the ownership, custody or location of a historical object ... The primary purpose of tracing the provenance of an object or entity is normally to provide contextual and circumstantial evidence for its original production or discovery, by establishing, as far as practicable, its later history, especially the sequences of its formal ownership, custody and places of storage. The practice has a particular value in helping authenticate objects.*" (Provenance [updated 2018])

In digital contexts it can also be applied to digital objects, digital representations of objects, data and metadata. In this case provenance can be recorded for physical objects, digital objects and data in the context of all kinds of collections and archives.

## Quality assurance

In order to promote reuse of metadata and objects the metadata must be correct, complete and up-to-date. To this end a practice of quality assurance for metadata must be adopted. This implies the establishment of requirements the metadata must comply to, monitoring these requirements, verification of the actual metadata, assigning sufficiently qualified staff. The quality requirements at least include the minimum requirements proposed here, with the addition of criteria for the specific domain.

## Roadmap to FAIR collections

---

In order to make collections FAIR, a LAM institution should draw up a roadmap, consisting of a number of steps. Some of these steps can be taken simultaneously, some are preconditions for other steps, some are temporary, others will be permanent.

We recommend the following steps:

### Set up a working group

The first step should be setting up a working group, taskforce, or similar, to coordinate

and carry out the activities needed for making the institutions collections FAIR. The working group should consist of representatives from all internal stakeholders and specialists, such as metadata experts, developers, systems coordinators, legal copyright experts, etc. The intention is to cover all perspectives of FAIRness: description, hardware, software and legal issues.

The working group could be established as a temporary task force until all collections are FAIR. But it is recommended to have some kind of permanent body to monitor and adjust the situation as needed.

## **Compile an overview of collections**

One of the first tasks of the working group is to compile an overview of all of the institution's collections, and all the databases, repositories and systems these collections are described and presented in.

## **Compile an overview of existing agreements**

An overview of agreements and collaborations of the institution with national and international standards organizations will be helpful in assessing what standards have already been adopted, implemented or agreed upon.

## **Compile an overview of existing FAIR principles per collection**

The overview of collections and the overview of agreements can serve as input for compiling an overview of already implemented FAIR principles per collection. Besides these overviews of course also thorough examination of the collection infrastructure is indispensable. A useful instrument would be a kind of scorecard containing all FAIR principles on which one can mark the score for each principle per collection, including comments. Obviously, the scores can only be established by communicating with all stakeholders. A completed scorecard for each collection will automatically reveal the issues that have to be addressed.

## **Involve stakeholders**

In order to identify the FAIR issues that your users experience, it is recommended to involve all stakeholder groups: end users, researchers, teachers, etc. An advisory group can be established with representatives of all stakeholder groups that are relevant to the institution. Communication between working group and advisory group can produce an overview of usage requirements, use cases, etc.

## **Talk to others**

If possible, talk to other institutions who already managed to make collections FAIR and learn from their experiences. There is no need to invent the wheel again.

## **Execute a pilot**

If appropriate, carrying out a pilot project with a sample collection or sub-collection can be useful to identify the main issues. The pilot can for instance focus on providing bulk downloads of collection data, connecting the collection to other platforms using linked open data, or publishing digitized objects. The pilot will reveal all obstacles that are preventing findability, accessibility, interoperability and reusability of the specific collection's data and objects.

## Establish a FAIR policy

An institutional policy should be established for achieving FAIR collections, preferably based on a pilot, feedback from stakeholders and the compiled overviews.

## Write an Implementation plan

Once an institutional FAIR policy is in place and solid overviews of collections, agreements and issues are available, it is time for writing an implementation plan. In this plan it is essential to set priorities and establish implementation decisions for those principles that are fundamental across all collections. For instance, in order to make anything FAIR, you need globally unique persistent identifiers. So before you do anything else, it is necessary to decide on the specific PID-schemes available (DOI, Handle, URI, etc.) that are suitable, identify the practical implementation issues involved and configure a PID infrastructure.

Other fundamental principles are sustainable storage, license management, links/reference management, provenance information and metadata quality.

The information system/database/software infrastructure is also important. An institution has to be able to implement FAIR principles in the available technical infrastructure. If the institution is dependent on external system and database providers, there have to be plans for negotiating with these providers if necessary. Also, when procuring new database/system environments, it is advisable to incorporate the FAIR principles into the requirements.

When these organizational and technical foundations are adequate, it is time to focus on the systems infrastructure. This depends highly on the degree of distribution and fragmentation of the systems infrastructure. A museum with just one information system for both cataloguing and discovery can focus on that one system only for implementing most of the technical requirements. A university library with physical text material, online digital resources, scholarly output and special collections will have a number of systems and databases for cataloguing, handling and presenting their collections. In this case focus on source systems/databases first, so all improvements can be propagated to dependent target systems. For instance, objects can be described in a backoffice library management system, the metadata can be copied to an end user discovery tool and a repository, the digital objects can be made available in an institutional repository and copied to an online end user tool.

Finally, it is essential that existing workflows and responsibilities are adapted to the new FAIR principles, so that new objects and collections are FAIR from the start.

## Conclusion

---

Making LAM collections FAIR (findable, accessible, interoperable and reusable) is a complex endeavor. It is important that all involved organizational units of an institution work closely together in order to break down the organizational, legal and technological barriers that prevent it. Besides that, external stakeholders, users of the collections and the collection data, have to be involved. As noted, this process requires addressing a wide range of facets when adopting FAIR principles. Therefore, the strategy must be methodical and systematic. Guidelines must be determined, and a policy must be established for the three collection levels “objects”, “metadata” and “metadata records” based on a pilot. It is important to observe, evaluate and, if necessary, adapt the guidelines with every new implementation. This approach should lead step by step to make an institution’s collections FAIR.

## References

---

- Annotations and FAIR [Internet]. [2018]; [cited 2018 Feb 24]. Available from: [https://docs.google.com/document/d/1UObmtNCL\\_Dw5\\_tQLJkWgfkBsm\\_4YpCXxiu3zj2Fsuv8/edit](https://docs.google.com/document/d/1UObmtNCL_Dw5_tQLJkWgfkBsm_4YpCXxiu3zj2Fsuv8/edit)
- APEF, CLARIN, DARIAH, Europeana, E-RIHS, Iperion-CH, PARTHENOS [Internet]. [2018] Cultural Heritage Data Reuse Charter: the Mission Statement; [cited 2018 Feb 24]. Available from: <https://sondages.inria.fr/index.php/593568/lang-en>
- Berners-Lee, T. [Internet]. [updated 2015 Aug 31]. 5-star Open Data; [cited 2018 Feb 24]. Available from: <http://5stardata.info>
- British Library Free Data Services [Internet]. [2018]; [cited 2018 Feb 24]. Available from: <http://www.bl.uk/bibliographic/datafree.html>
- CC0 [Internet]. [2018]; [cited 2018 May 1]. Available from: <https://wiki.creativecommons.org/wiki/CC0>
- CLARIAH [Internet]. [updated 2017 Aug 9]. Cultural Heritage Data Reuse Charter : Mission Statement. Amsterdam: CLARIAH; [cited 2018 Feb 24]. Available from: <https://www.clariah.nl/en/new/news/cultural-heritage-data-reuse-charter-mission-statement>
- Creative Commons. Attribution 4.0 International [Internet]. [2018]; [cited 2018 Feb 24]. Available from: <https://creativecommons.org/licenses/by/4.0/>
- DEN [Internet]. [updated 2016 Nov 23]. DE BASIS. DEN; [cited 2018 Feb 24]. Available from: <http://www.den.nl/debasis>
- The FAIR data principles [Internet]. [2014] La Jolla, CA: Force11; [cited 2018 Feb 24]. Available from: <https://www.force11.org/group/fairgroup/fairprinciples>
- Getty Research Institute [Internet]. [updated 2017 Mar 7]. Art & Architecture Thesaurus® Online. Getty Research Institute; [cited 2018 Feb 24]. Available from: <http://www.getty.edu/research/tools/vocabularies/aat/>
- Hakala J, Network Working Group [Internet]. [updated 2001 Oct]. Using National Bibliography Numbers as Uniform Resource Names. Request for Comments. The Internet Society; [cited 2018 Feb 24]. Available from: <http://www.ietf.org/rfc/rfc3188.txt>
- ISO [Internet]. [updated 2012 Sep] ISO 14721:2012 (CCSDSS 650.0-P-1.1) Space data and information transfer systems — Open archival information system (OAIS) — Reference model. Geneva: ISO; [cited 2018 Feb 24]. Available from: <https://www.iso.org/standard/57284.html>
- Library of Congress. Subject Headings [Internet]. [2018]. Library of Congress; [cited 2018 Feb 24]. Available from: <http://id.loc.gov/authorities/subjects.html>
- Metadata2020 Librarians [Internet]. [2018]; [cited 2018 Feb 24]. Available from: <http://www.metadata2020.org/communities/librarians/>
- Mons B, Neylon B, Velterop J, Dumontier M, da Silva Santos, LOB, Wilkinson, MD. 2017.
- Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use [Internet]. [cited 2018 Feb 24]; 37(1):49-56. Available from: <http://doi.org/10.3233/ISU-170824>
- Martone M. 2017. Annotating All Knowledge, FAIRly. Hypothes.is [Internet]; [cited 2018 Feb 24]. Available from: <https://web.hypothes.is/blog/annotating-all-knowledge-fairly/>

National Center for Biotechnology Information. MeSH [Internet]. [2018]. Bethesda MD: National Center for Biotechnology Information; [cited 2018 Feb 24]. Available from: <https://www.ncbi.nlm.nih.gov/mesh>

Open Definition 2.1 [Internet]. [2018]. Open Knowledge International; [cited 2018 Feb 24]. Available from: <http://opendefinition.org/od/2.1/en/>

PARTHENOS/About the project/the goal [Internet]. [2018]; [cited 2018 Feb 24]. Available from: <http://www.parthenos-project.eu/about-the-project-2/>

PARTHENOS. 2017. Report on Guidelines for Common Policies Implementation (1). D3.1. [Internet]; [cited 2018 Feb 24]. Available from: [http://www.parthenos-project.eu/Download/Deliverables/D3.1\\_Guidelines\\_for\\_Common\\_Policies\\_Implementation.pdf](http://www.parthenos-project.eu/Download/Deliverables/D3.1_Guidelines_for_Common_Policies_Implementation.pdf)

Provenance [Internet]. [updated 2018 Jan 19]; [cited 2018 Feb 24]. Available from: <https://en.wikipedia.org/wiki/Provenance>

Recommended formats statement [Internet]. [2016] Library of Congress; [cited 2018 Feb 24]. Available from: <https://www.loc.gov/preservation/resources/rfs/>

Rijksmuseum API [Internet]. [2018]; [cited 2018 Feb 24]. Available from: <https://www.rijksmuseum.nl/en/api>

Universiteitsbibliotheek Gent Open Data [Internet]. [updated 2018]; [cited 2018 Feb 24]. Available from: <https://lib.ugent.be/en/info/open>

What is open? [Internet]. [2018] Open Knowledge International; [cited 2018 Feb 24]. Available from: <https://okfn.org/opendata/>

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons, B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [Internet]. [cited 2018 Feb 24]; 3:160018 Available from: <http://doi.org/10.1038/sdata.2016.18>

## Notes

---

[1] See also “The library as an open global platform”, <https://future-of-libraries.mit.edu/>

[2] The article is based on an internal advice for the Library of the University of Amsterdam by the authors (2018)

[3] The Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

[4] Articles 19-21 Dutch Copyright Act

[5] In English “Digital Heritage – Building a Successful ICT Strategy.”

[6] Patent and trademark rights are not licensed under a Creative Commons license.

## About the authors

---

Lukas Koster MSc (l.koster@uva.nl) is Library Systems Coordinator at the Library of the University of Amsterdam and the Amsterdam University of Applied Sciences (HvA), focusing on data infrastructure and discovery. His current activities are project manager Improvement of the functionality and user experience of the Primo Discovery

tool, co-lead of the Working group FAIR Collections and project team member of the Amsterdam Cultural Heritage Linked Open Data Network project. ORCID: 0000-0003-0214-4721.

Saskia Woutersen-Windhower LLM ([s.windhouwer@uva.nl](mailto:s.windhouwer@uva.nl)) is specialist electronic publishing/repository manager and open access service manager at the Library of the University of Amsterdam (UvA) and the Amsterdam University of Applied Sciences (HvA). She is also involved in Open Science at the UvA and in the Horizon2020 project OpenUP, and co-lead of the Working group FAIR Collections at the UvA. ORCID: 0000-0003-0120-266X.


Subscribe to comments: [For this article](#) | [For all articles](#)

### Leave a Reply

Name (required)

Mail (will not be published) (required)

Website

I'm not a robot   
reCAPTCHA  
[Privacy](#) - [Terms](#)

---

[Log in](#)

This work is licensed under a [Creative Commons Attribution 3.0 United States License](#).

