Best Practices for Textile Collections about Documentation & Digital Data Curation



November 2021

1 Te: Mo	xtile Museums Are oving into the Digital Era	P.3
2 Pr	oritize Access Over Quality	P. 5
3 Ma No	nage People and IPR, t Just Data	P. 8
4 ^{Op} Cu	en Access Repositories for ltural Heritage Data	P. 11
5 Sco Tra	enarios for Digital ansformation.	P. 13
6 Ge	tting into Specifics	P. 18
7 ou	r Top 5 Recommendations	P. 25
8 Th	ank you!	P. 27

The following guidelines have been prepared, first and foremost, with small and medium-size textile museums in mind. Of course, large national museums may also benefit from them, although their needs and resources tend to be quite different.

Textile Museums Are Moving into the Digital Era

Anybody else, out there?

For some people, collections of historical textiles might seem too small a niche, but facts prove this to be wrong. For instance, take a look at We Wear Culture, from Google Arts and Culture, or Europeana Fashion. On the more specialized side, some interesting projects are making headway. Feel free to check them out for inspiration and stimulus.

SilkMemory provides access to the archive database of the Lucerne School of Art and Design with digitised text and image sources about the silk industry of the Canton of Zurich. Born after the commercial demise of the once-thriving Swiss silk industry, it provides a thoughtful answer to a danger that is common to many European countries: the dispersal and loss of the valuable archival and material heritage generated by those industries. It offers a database of fabrics, books and images kept in their archives.

>> https://silkmemory.ch

The **ART-CHERIE** project brought together partners from Belgium, Greece, Italy and the United Kingdom, from a quite a broad scope, including an interesting connection to the training of fashion designers. Among other outputs, it aimed at providing a Digital Database of Museo del Tessuto di Prato Exhibits and Collection". However, it is not openly accessible.

>> https://www.artcherie.eu

IMATEX is the online database offering information about the collection of the Centre de Documentació i Museu Tèxtil de Terrassa. Created in 1996, it was originally built as a gateway for designers searching for inspiration in CDMT's historical collection, and later transformed into a generic online information resource. It is extremely rich in content, including costumes, accessories, designs, paraments, sample books, a library and an outstanding collection of more than 9,000 textiles.

>> https://www.inmatex.cdmt.es

The **MINGEI** project aims at exploring the possibilities of representing and making accessible both tangible and intangible aspects of craft as cultural heritage. One among the crafts under study is silk weaving, led by one of the project partners, Haus der Seidenkultur in Krefeld. The project does not directly intend to build a database, but rather a repository of innovative storytelling models, including interactive Augmented Reality and Mixed Reality. It does have a strong emphasis on developing content description tools that comply with existing semantic web standards, such as CIDOC-CRM.

>> https://www.mingei-project.eu

The **PARVENUE** project, led by art historians from the Heinrich-Heine-Universität in Düsseldorf, is built on a tight collaboration with the Deutsche Textilmuseum in Krefeld, one of the European capitals of silk industry. One of the areas of the project is built on a preliminary cataloguing of the collection of 30,000 fabrics and costumes in the museum, not yet available online.

>> https://www.parvenue-projekt.de

The **Restaging Fashion** project purports to build an online catalog of costumes, prints and drawings, held in different German institutions, adding 3D visualizations of some of these objects.

>> https://uclab.fh-potsdam.de/projects/restaging-fashion

Of course, **SILKNOW** (the research project that has produced this very document) should also be a source of useful information for anyone interested in historical fabrics. It has a very broad set of goals, one of them being an enhanced access to the information about European silk heritage that is scattered across many institutions worldwide. ADASilk, our exploratory search engine giving access to collections from dozens of museums, is a proof of concept on many of the ideas and approaches outlined in this document. Our publications can also provide information and guidance.

>> https://silknow.eu

Prioritize Access Over Quality



Should my database be perfect? Should it? Should I?

This sounds scary for most of us working in the humanities. We love details and feel insecure about providing information that is not in accordance with high academic standards. As explained below, in the current situation, this is more a hindrance than a danger, and we suggest beginning -or continuing- to publish your data as much and as openly as you can.

If you think your database is a mess, you are not alone. Just having your data on an Excel file is much better than the average situation. Things happen along the way, sometimes for good, sometimes not. However, it is better to begin with something, however inadequate, than to wait for the ultimate, authoritative catalogue.

- "We are looking forward to reviewing our catalogue". Well then, welcome to the boat! This is the permanent situation. Cataloguing is always a work in progress.
- Prioritise making images and a small bit of textual information available (if that is all you have), even if they are far from ideal. If you can have professionally made photographs, that's great. If you only have old or uneven quality ones, go ahead and publish them, that's better than nothing.

Establish and follow some workflow priorities, for your online catalogue publishing. For instance, publish first what is on show (permanently or in temporary exhibitions); second, loans and new accessions; and last, all the rest... Trying to do all at once will not work.

 Another common approach is to bear in mind three possible kinds of pieces in your collection: top-notch, second-tier, and (a huge amount of) minor pieces. It makes sense to treat their information in different ways within your database: for instance, providing much more information about the former than the latter. At least, publish all that you exhibit; it will be good enough, usually.

Don't blame yourself about publishing information that you feel insecure about. Since your cataloguing is not perfect, and never will, let the users know about it. For instance, with a general disclaimer in the search interface, or with reminders for specific parts of your data. Accept imperfection and ask for understanding. It's better than hiding the collections for years, decades, or... Be open to feedback and disagreement from the users. They may provide useful corrections. Researchers will contact you, asking for help or giving advice. Academia will knock at your door, and that should be fine. Non-experts may want to reach the museum, too, and they have a right to do so; quite often, that heritage belongs to them, as well. Provide simple, specific ways to interact with information on the catalogue. Most times, an email address for comments will be good enough. (Of course, then make sure that someone reads and answers the comments...).

Terminology standardization is key, so please try to follow it as much as you can. Even more so, in a digital world, where each institution should share its information and not be a self-contained silo. Some degree of information standardization is necessary: terminology, information standards, platforms, mapping between data models... Be ready to advocate for it (more on this below). However, understand that not everyone cares equally about standardization, which makes them no less of a good professional. It makes sense, since many times textile experts are exactly that, specialists. They care about the specific.

For the same reason, always bear in mind that all cultural heritage is, by essence, fuzzy and complex. When dealing with information about your objects, **embrace ambiguity:** polysemic terms, uncertain dates, alternative labels, changing attributions, evolving standards... (This can be hard to accept for ICT collaborators, but there's no way around it.) Cataloguing rules and standards allow to handle this uncertainty, to a small degree. Nevertheless, when exporting and sharing that information, many of those details can get lost or become less clearly retrievable. Moreover, while using accurate language is necessary, not everyone is an expert in textiles. Lots of people love textiles and fashion but completely ignore its specialized terminology, since it is a very technical field. When possible, provide information that is not just accurate, but conceptually accessible for non-specialized audiences, too. For instance, with explanations or illustrations that clarify a technical term, or with alternative, simpler terms. Be reminded, too, that language and people evolve over time.

Here you are with a list of thesauri and vocabularies that might help you when documenting and cataloguing your records:

The SILKNOW Thesaurus is dedicated to the specific vocabulary of historic silk textiles, which also includes local term variants. It collects more than 800 terms that not only come from academic and historical sources but also from oral history, as knowledge coming from weavers is something that SILK-NOW intends to preserve. Terms include silk weaving techniques, designs, looms, costumes, etc. All of them in four languages: Spanish, French, English and Italian.

>> https://skosmos.silknow.org/thesaurus/en/

The Textile Museum Thesaurus is based on the collection of the George Washington University Museum, meaning that only terms in use in the museum's catalog records are included within these hierarchies. It functions both as a controlled vocabulary for cataloging their collections and as an electronic querying aid, structured to permit the highest level of success in retrieving the appropriate objects during searches.

>> https://museum.gwu.edu/textile-scholarship

 The Europeana Fashion Vocabulary is built in 10 languages (English-French-German-Dutch-Italian-Serbian-Swedish-Spanish-Portugese-Greek) for fashion and fashion-related concepts. It focuses not only on the objects that are the result of creative processes in the fashion industry, but also on the processes themselves and the resources and techniques used (with a focus on object – material – technique).

>> https://cordis.europa.eu/docs/projects/ cnect/7/297167/080/deliverables/002-EuropeanaFashionDeliverable23EuropeanaFashionThesaurusv1.pdf

 ICOM Vocabulary of Basic Terms for Cataloguing Costume. It offers comprehensive guidance on cataloguing costume to ensure that the information on each garment in museum collections is recorded clearly. The lists of basic terms for women's, men's, and infants' garments have been published as a guide, setting out what information should be recorded, not just about garments and accessories dresses, but also about the many aspects of its study.

> >> http://terminology.collectionstrust.org.uk/ ICOM-costume/

Vocabularies from CIETA, the Centre International d'Étude des Textiles Anciens. Since CIETA's foundation, its members have sought to establish a precise terminology for the tools, processes and structures of weaving. Therefore, they have published a number of vocabularies, each focusing on one particular language, with the definitions of the terms listed and their equivalents in a set of other languages.

>> https://cieta.fr/cieta-vocabulaire/

The Getty Art & Architecture Thesaurus (AAT). It includes generic terms, associated dates, relationships, and other information about concepts related to or required to catalog, discover, and retrieve information about art, architecture, and other visual cultural heritage. This include related disciplines dealing with visual works, such as archaeology and conservation. Their goal is to be ever more inclusive of various cultures and their visual works. Also, in recognition of diverse collections found in art museums, the AAT contains terminology to describe objects and associated activities that are ceremonial or utilitarian in nature but are not necessarily labeled as art according to traditional Western aesthetics.

>>http://www.getty.edu/research/tools/vocabularies/aat

On the other hand, your catalog should reflect the huge diversity of textiles, and their terminology. Be open to synonyms, including local and popular usage.

Standards and cataloguing rules exist and are sometimes officially adopted, but then, cataloguers don't always follow them properly! Train your staff and support them, ensuring a proper use. Otherwise, there is no point in adopting any standard.

It is never easy to strike the right balance between quantity and quality of information. Having said all of the above, in our current environment (pandemic, digital transformation...), **quantity** is more important than **quality**. In our opinion, and for the time being, scaling up, offering more data, seems more of a priority. In the future, however, it will be necessary to focus on quality, as a necessary correction.



Manage People and IPR, Not Just Data

Documenting the pieces in a collection may not be the most exciting aspect of our work, but it is a core part of it, as well as a legal obligation, in many cases. Nowadays, moreover, digital information is a key driver for our society, with many new possibilities to share it across institutions.

* Cataloguing your collection might be a life saver (and a heritage saver), specially in case of natural or human disasters. This task is important enough to have **dedicated staff**. You don't need to have a full-time specialist for your inventory or catalogue, but the responsibility should be clearly incorporated within the museum structure.

* Don't throw away old information. Don't forget that today's cataloguing will also become obsolete. The **memory of the museum** is part of its mission, too. For instance, treasure your paper-based records. Digitize your archival information, and make it accessible.

✓ Talk to your supervisors and ask them to decide who is in charge of the management of digital information within the institution. (It might be a shared assignment.) It is an essential task for any heritage collection, and responsibility over it should be clearly established and supported by decision makers and budget planners.

✓ Sharing digital information with external repositories should not mean extra work only for the museum staff. Talk to the staff in charge of the repository to find a workflow that accommodates both sides.

✓ Not all information is equally useful, so feel free to select what seems more worthy of attention for your users and prioritise it. Remember that some exceptions to open access should be kept in mind. Not just copyright-protected materials, but also objects from some traditional cultures. Of course, there is no need to open up all your information. Administrative details are nobody's business, outside the museum.

* There is a wide range of intellectual property licences available. It's not just either public domain or "all rights reserved". Creative Commons licences provide many layered, nuanced variants, fit for most cultural institutions.

* Some copyright holders may not necessarily require economic compensation but at least recognition and respect to their cultural identity should be secured and made public.



Figure 1. CC License Freedom Scale Chart, by Foter. CC BY-SA.

* Institutional guardianship, recognition of authorship or any other moral rights, however, must always be respected. Any **rights** associated with the object or its derivative data must be recognized and made visible to users. Again, there are many options available in open-access licenses that should be enough to accommodate the needs of most heritage institutions, including the smaller ones.

works are allowed but must not be shared.

✓ Understand the way your collection / institution was built. On this issue, museums specialized in textiles and those with broader, more diverse collections, are quite different. The same applies to size: small museums may learn from the biggest players, but you don't need to imitate a different kind of museum. Textile collections are also very diverse, in themselves: some may be focused just on fabrics, others on fashion, others are old factories with historical looms and tools... Catalogue records should respect and record that diversity.

∼ Be smart with external collaborations. Negotiate your conditions. For instance, you can treat profit-making partners differently from non-profits or public authorities. Demand recognition and, where appropriate, returns for your work. They need not be only monetary: sponsorship, collaborations, technological partnerships, capacity-building, networking, reputation... can also be valuable sources of support for cultural institutions.

✓ Small and medium size museums or collections can benefit greatly from alliances with local authorities, universities, and companies, in applying for external funds for their cataloguing, digital data management, implementing open-access repositories and platforms, etc. The EU and other funding bodies are very interested in projects that bridge digital technologies and cultural heritage, and they expect those projects to be multi-actor and interdisciplinary. ✓ In many cases, authorship of records or metadata is not properly recorded. It's a matter of practice, not of standards, since most of them indicate ways to include that information. We recommend the general publication of this information, something that can be done with a layered approach, depending on the provided depth of details.

✓ Income from reproduction fees is rarely worth its own costs, in many institutions, let alone a profitable source of funding. Big players among museums are leading a path towards making collections fully open-access, by providing high quality information, free of charge. This will increasingly become the standard for most institutions, be it de facto or by legal requirement. It makes sense to keep it in your horizon and work towards that framework.

✓ For the largest part of our historical collections, that predate the 20th century, copyright on the original objects is not an issue. In a great majority of cases, there is no legal basis to limit access and usage to the information about them.



Open Access Repositories for Cultural Heritage Data

Q Peers and colleagues, not just the general public, will particularly appreciate it if you **share as much data as possible** about your holdings. It will also allow you to receive comments, help and proposals from them, in most cases. It's a win-win!

Online publishing of your collection information is no easy task, so make sure to get expert advice from trusted colleagues or your software providers. Free alternatives, like Omeka, are a simple way to begin, but as you begin to ask more from your system, commercial options will become more interesting. In an ideal setting, it should be possible to access your information repository through both a human-readable interface and an application programming interface (or API), one that connects computers to each other. (If still wondering what that means, check https://medium.com/open-glam/backfrom-the-first-api-culture-day-part-2-e9547bd08dea for an intro.) This is not (yet) essential, but at some point it will.

Use whatever software / platform / repository / collections management system you like or can afford, but please make sure that it can **export** your data to other formats. (A good old .csv file is a suitable start for most needs.)

Try to keep everything in the same database and show the information to different kinds of users in different ways, according to their profiles. It will make the management of that data much simpler and safer.

As far as possible, adopt a **structured format** (data model or standard) for your information. Take into account that the fields in your records will have to be mapped to those in the destination repository or database.

Don't panic. There's a growing community supporting open access efforts within the cultural heritage sector.

 OpenGLAM is an excellent starting point, and a great source of information.

>> https://openglam.org/

If you are interested in contributing data to Wikimedia Commons (the repository that feeds many open-access resources, Wikipedia among them), check the GLAMWiki:

> >> https://outreach.wikimedia.org/wiki/GLAM/Get_ started

Europeana is the largest open-access repository for European cultural heritage of all kinds. Collaborating with them will be a huge step to improve your information. They do not work directly with data providers (i.e., cultural institutions willing to contribute with their data like, for instance, your museum). Instead, they gather data through national or thematic aggregators. These are the ones you should reach. Europeana also provides a strong network of experts working with digital cultural heritage.

>> https://pro.europeana.eu/page/aggregators https://pro.europeana.eu/network-association/sign-up

¬ Specifically for fashion, EFHA − the European Fashion Heritage Association is a terrific resource (and a Europeana aggregator, by the way).

>> https://fashionheritage.eu/

■ For historical textiles, mostly those made of silk, but not only, the SILKNOW team is the people to go to (ok, full disclosure, that's us, but you get it).

>> http://silknow.eu

The Programming Historian is a journal offering novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows. Even more interestingly, it's a volunteer, community-driven resource.

>> https://programminghistorian.org/en/lessons/

O Visibility of the owning institution must always be provided by the external repository. If technically possible, a permanent link to the same object in the museum online catalogue, should always be available, thus giving users the possibility to see the original record and to contact the institution responsible for its conservation, study and dissemination.

O Mindset changes take time. You may want to use the following **strategies**:

People and institutions evolve. Even in museums. But we all need help and recognition. Set small goals, provide guidance and support, appreciate efforts, and make success stories known internally and externally.

■ Invoking "digital transformation" might be useful for in-house discussions on all these issues. The term itself is contested, as usual, but we like the definition from Public Digital: "digital transformation is the act of radically changing how your organisation works, so that it can survive and thrive in the internet era". How's that? Will it work for you and your colleagues? We think so!

5 Scenarios for Digital Transformation

Scenario A – The Good Old Times

"All of our information on the collection is stored on paper – yes, only on paper. We've been thinking about going digital but don't know how to succeed, not even where to start."

>> Some ideas to consider <<

Build digital skills for you and your staff. This can take many forms, from taking a graduate degree to attending short online courses, or even independent training, but digital transformation will not happen through goodwill only. Upskilling the staff in charge is a must.

Learn from your peers. Find museums with a similar profile, that are already digitally-savvy. Ask them about their mistakes and their successes (not necessarily in that order). Professional associations, conferences, workshops, etc. are usually the place to start.

Many companies are specialized in digitizing information. Relieving your staff from that additional burden will usually be the way to go, especially if there is a lot of data to handle and you are already very late for the information age. It might be quite an investment but grants and other sources of public funding for these processes are more and more common, nowadays. Get ready for external funding applications.

Before acquiring any software, you should consider issues such as who will manage the database, or the extent of your financial resources to buy a CMS? See the Getting into Specifics section below, for a listing of available software, including some free, open-source options.

Keep your users in mind from day zero. We tend to think first about museum staff and then imagine some kind of general audience that will happily read whatever we give them. Sadly, it never works that way. Do serious research about the profiles of your users, beginning with your staff and then proceeding to visitors. Look into your current and your desired audiences, their needs, wishes, expectations... and then plan your information around those profiles. This will save you many later disappointments (like "why does our website have such poor traffic stats?"). First, make a pilot study, digitising just a small cross-section of your collection, and evaluate the results. With that experience in mind, organize a digitization calendar or campaign, and go ahead.

Solution Once you're done and have a new shiny database, please keep your paper records! They are an essential part of your institution's memory.

>> Practical guidelines and best practices <<

This Small Museums Cataloguing Manual introduces users to the cataloguing process, from planning and procedures to digitisation and management of digital collections: https://mgnsw.org.au/ sector/resources/online-resources/collection-management/ collection-management-systems/

The path of successful digital transformation has been walked by many small museums by now. The Swedish LSH museum provides a very informative and realistic example: https://pro.europeana.eu/ post/making-impact-on-a-small-budget

In case you want or have to deal with your records on paper on your own, OCR (Optical Character Recognition) will be part of your everyday work. The following tutorials can help you to get along:

- Generating an Ordered Data Set from an OCR Text File.

>> https://programminghistorian.org/en/lessons/generatingan-ordered-data-set-from-an-OCR-text-file

- Cleaning OCR'd text with Regular Expressions.

>> https://programminghistorian.org/en/lessons/ cleaning-ocrd-text-with-regular-expressions

- OCR and Machine Translation.

>> https://programminghistorian.org/en/lessons/ OCR-and-Machine-Translation

- Detecting Text Reuse with Passim.

>> https://programminghistorian.org/en/lessons/ detecting-text-reuse-with-passim

- Cleaning Data with OpenRefine.

>> https://programminghistorian.org/en/lessons/ cleaning-data-with-openrefine

Scenario B – Perfect Digital Storm

"Yes, we're mostly digital, but... Some parts of our collection are already catalogued on digital support, but very unevenly: different databases, odd cataloguing standards and rules, little use of controlled vocabularies... Some of that information is outdated, questionable, or plain wrong. Of course, small sections are not yet catalogued. Can't imagine even trying to put all that on our website."

>> Some ideas to consider <<

It's ok to establish priorities within your collection. (In fact, please do so.) Sometimes, specific sections of your textile holdings simply do not require a detailed, piece-by-piece cataloguing. Yes, it would be fun to have it, but it would just take too much time, and there isn't such a great demand for them. They might be interesting only for specialists, or still lack basic supporting research, or be too extensive and repetitive... Basic cataloguing, one that just indicates the existence of those pieces (or groupings of them) and a few simple pieces of data, is a very reasonable option in those cases.

At the same time, set a calendar of goals to fill those gaps in your information. Try to coordinate your efforts and obtain additional help: a research project, a conservation campaign, a temporary exhibition, a loan request, a new fundraiser event... These and other occasions should serve as stimuli and sources of support for a better coverage of your collection.

Patchy, uneven information is better than no information at all. The more you give your users, the more you will get from them. Waiting to go online until you have your perfect catalog will take forever. Begin publishing some information, even if you are not entirely happy with it, and tell users that you know it's imperfect. Ask for understanding, and help.

Based on your experience, try to understand which tasks belong to your core mission. It's best to cater to those with your permanent staff. Other tasks can be outsourced more efficiently, pooled and shared with other museums, collections, etc. In this regard, documenting and sharing information on your collection does belong at the center of your institution, in our opinion. Build it into your main responsibilities, and make sure that decision makers support it with appropriate resources.

>> Practical guidelines and best practices <<

Cataloguing historical textiles is an art in itself. The closest thing to a cataloguing standard is that provided by CIETA, but it isn't easily accessible, and can be too demanding for most no-specialized catalogers. If you have the training and the resources, use it. Otherwise, feel free to establish you own in-house standard, one that suits your collection and your resources. In any case, do not invent the wheel again: look at museums whose documentation you know and admire, and learn from their records.

Work with a single yet flexible database or CMS, as those listed below in section 6.b. You will always have to find a compromise between standardization and the diversity or degree of detail of your information. Approaches will necessarily differ between small, specialized collections and larger, more heterogeneous museums. One size does not fit all.

Using standard terminology, and doing it right, is no easy task. It takes time and effort, but results always pay back. Luckily, there's already a great deal of progress in this subject: check the thesauri in section 2.

Scenario C – Already Online but Not Yet Open

"We have an online catalog that offers the highlights of our collection. The thing is, we're sharing just a tiny part of the information we have. We would love to provide open access to our data but are unsure on how to do it properly."

>> Some ideas to consider <<

Be ready to advocate for open access within your institution. There's still a lot of traditional prejudice (and some reasonable precaution) to be found in this regard.

Set a clear policy about IPR on your contents, including information about your collections, both text and images. Let users know what they can or can't do with them. Give due recognition to all parties involved: authors, compilers, etc. Make sure that the same happens when your information is available through external repositories.

Most authorities are giving full support to open access efforts, in the understanding that public funding should benefit everyone. There might be legal obligations or grant requirements that you must fulfill, now or in the near future. Quite likely, this is going to become standard for many heritage institutions. Get up to speed on it.

Integrate online and on-site experience. It usually helps everyone understand the importance of open access and semantic web technologies. For instance, extending on-site experience (via a visitors app, or a version of your website) with information about other pieces that are not on view, or related pieces in other locations, or additional information – current or past events, publications, audiovisual materials, etc.

Making your data interoperable can be a boon or a bane. It will depend on many circumstances. Standards, data models, mappings, licenses, file formats... ok, it looks daunting, but if your datasets are reasonably well organized and homogeneous, sharing them through external repositories (like Europeana or Wikimedia Commons) should not be overly difficult.

Join the open access community. There's much to learn from people who are already working in this area, and quite often they are willing to provide guidance.

>> Practical guidelines and best practices <<

Intellectual property is key, when working with shared repositories. You will need to have clear policies and agreements. The following resources will provide you with good examples and some templates, but your collection's legal officers will definitely want to make sure everything fits their requirements.

RightsStatements.org currently provides 12 different, standardized terms or 'rights statements' that summarize the copyright status of objects in cultural heritage collections, as well as how those objects may be used.

>> https://rightsstatements.org/en/

The Europeana Data Exchange Agreement (DEA) structures the relationship between Europeana and its data providers.

> >> https://pro.europeana.eu/page/ the-data-exchange-agreement

 For Creative Commons licenses, check section 3 above.

Making your data and the repository interoperable will require you to tweak with your metadata, mapping your data structure to theirs. A model case is Europeana, having already a lot of experience and resources about these processes:

https://pro.europeana.eu/share-your-data/metadata

Their guidelines give detailed support for providers wanting to map their data to the Europeana Data Model.

>> https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf

Many mappings, refinements and extensions are already available, based on work by Europeana and partners that adapt EDM to the needs of specific domains and applications or make systems that use EDM interoperable with other data models.

>> https://pro.europeana.eu/page/edm-profiles

CIDOC CRM is the standard data model for museums and cultural heritage institutions. It is highly complex yet very flexible, allowing the accommodation of almost any kind of pre-existing data structure. Many institutions have already followed these mapping procedures, so there's a reasonable amount of guidance to learn from, as well as one tool available for use.

> >> https://cidoc-crm.org/sites/default/files/ Mapping_7_4_2003.ppt

>> https://cidoc-crm.org/mapping-tools >> https://cidoc-crm.org/reports_mappings

ContoME is an application designed to align the fields in your database to CIDOC CRM, or simply browse through the existing ontologies. This provides easily compatible models and access to robust interoperability.

https://ontome.net/



Getting into Specifics

a. Working Methods

Small Museums Cataloguing Manual, introduces users to the cataloguing process, from planning and procedures to digitisation and management of digital collections. Available for download in pdf format.

>> https://mgnsw.org.au/sector/resources/online-resources/collection-management/small-museums-cataloguing-manual/

Europeana's step-by-step process for museums and cultural heritage organizations provides guidance about sharing information through this common repository. It is very useful, not just for its specific purpose, but for any institution planning to improve their management of digital information about heritage collections. https://pro.europeana.eu/share-your-data/process Regarding IPR, Europeana also offers valuable guidance about assigning rights statements to cultural heritage content. https://pro.europeana.eu/ share-your-data/copyright

The Wikimedia Commons workflow offers similar guidance for data and media partnerships between Wikimedians and cultural institutions. The following overview is meant to help globally understand the overall workflow and to give pointers to the most often-used tools.

>> https://outreach.wikimedia.org/wiki/GLAM/Resources/ Data_and_media_partnerships_workflow

It is part of the general resources' wiki created by GLAMWiki to help GLAMs and interested Wikimedians:

>> https://outreach.wikimedia.org/wiki/GLAM/Resources

 Useful model projects and case studies, about both sharing digital collections and sharing knowledge, are available at

> >> https://outreach.wikimedia.org/wiki/GLAM/ Model_projects

The Programming Historian publishes novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching. All tutorials are rigorously peer reviewed to ensure the lesson works as intended and that all concepts are explained for a non-specialist reader. Some of them are particularly useful when transforming, filtering, or cleaning up data formats often used in the museum environment:

- Geocoding Historical Data using QGIS.
 >> https://programminghistorian.org/en/lessons/ geocoding-ggis
- Georeferencing in QGIS 2.0.
 > https://programminghistorian.org/en/lessons/ georeferencing-qgis
- Transforming Data for Reuse and Re-publication with XML and XSL.

>> https://programminghistorian.org/en/lessons/ transforming-xml-with-xsl

 Reshaping JSON with jq.
 > https://programminghistorian.org/en/lessons/jsonand-jq

A set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable (FAIR) were developed to facilitate good data management, which in turn leads to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. The FAIR principles were published in 2016. https://www.nature. com/articles/sdata201618

The Parthenos Guidelines to FAIRify data management and make data reusable offer a series of guidelines to align the efforts of data producers, data archivists and data users in humanities and social sciences to make research data as reusable as possible.

>>https://training.parthenos-project.eu/wp-content/ uploads/2019/02/20181219_FolderParthenos_FAIRifyDataManagement_PDFonly.pdf

b. Some CMS (Collections Management Systems) Designed for Museums

Spectrum: a UK-born collection management standard that is also widely used around the world. It offers a broad range of procedures for collections management for museums of any size and any collection type. It is offered in English, Arabic, Spanish, French, Dutch, Norwegian, Polish and Swedish.

>> https://collectionstrust.org.uk/spectrum/

Coeli is a cloud-based solution for cataloguing, managing and disseminating cultural heritage collections in an agile and intuitive way. Coeli includes a complete procedure-based collections management system together with powerful data harvesting and sharing capabilities.

>> https://www.coeli.cat/en/

TMS Collections and eMuseum: Gallery Systems' core collections management product, TMS Collections, is a browser-based application designed specifically for collections, content, media, exhibition, and loan management. Developed in partnership with museum professionals, TMS Collections can organize and manage all collection types.

>> https://www.gallerysystems.com/solutions/ collections-management/

MuseumPlus is a browser-based, comprehensive and globally used CMS developed by the Swiss zetcom Group. It offers state-of-the-art technology combined with a highly configurable frontend.

>> https://www.zetcom.com/en/museumplus-en/

Micromusée is a collections management system for all types of collections and museums. Version 7 offers a renowned interface for a better user experience and higher efficiency. It enhances the possibilities for shared or multilingual databases. Micromusée is user friendly and easy to customize.

>> https://www.axiell.com/fr/solutions/product/ micromusee/

OMEKA is a free, flexible, and open source web-publishing platform for the display of library, museum, archives, and scholarly collections and exhibitions. As a web application, it allows users to publish and exhibit cultural heritage objects, and extend its functionality with themes and plugins.

>> https://omeka.org/

Dédalo: open-source software platform developed by a small community of enthusiasts supported by a small company. Dédalo is freely available for individuals, organisations, enterprises or institutions to install, configure, and extend in accordance with their particular needs and without restrictions. It offers the possibility of a web publication API and a thesaurus.

>> https://dedalo.dev/

Matriz: it is the reference system for the inventory, management and online publication of cultural (movable, immovable and intangible) and natural heritage of the Portuguese ministry of culture. As an integrated information system, it includes components for automatic inventory management and online publication on MatrizWeb and also on the Europeana portal.

>> http://www.matriz.dgpc.pt/es_matriz30.php

eHive is a web-based collection cataloguing system. It is used worldwide by hundreds of museums, societies and private collectors to catalogue objects, store images, manage acquisition information and publish their collections online.

>> https://info.ehive.com/

S-Museum: it can be used in every type of establishment, from the smallest to the largest, for any kind of collection: art and history, archaeology, modern art, natural history. It is available in various languages and offers an open and interoperable system enabling data harvesting, data import as well as the availability of data and digital assets for third-party systems, websites, mobile applications, etc.

> >> https://www.skinsoft-lab.com/software/ museum-collection-management

More references at:

>> https://collectionstrust.org.uk/software/

c. Controlled Vocabularies

Please note that thesauri about textiles and fashion are already included in section 2 of this document.

Gazetteers of geographical names:

Geonames: a user editable geographical database available and accessible through various web services, under a Creative Commons attribution license. The GeoNames dataset differs from, but includes data from the US Government's similarly named GEOnet Names Server which establishes the official repository of standard spellings of all non-US geographic names.

>> http://www.geonames.org/

Pelagios: the Pelagios Network is a long-running initiative that links information online through common references to places. To create and maintain these connections, Pelagios has developed a method for creating semantic annotations, based on the W3C Web Annotation standard; tools and specifications for creating and making use of these annotations, most notably Recogito, an open-source platform for geo-annotating texts, images and databases; a community of individuals and organizations working with geographic data in humanities disciplines (history, language and literary studies, archaeology, etc.), and cultural heritage (galleries, libraries, archives and museums).

>> https://pelagios.org/

The Getty Thesaurus of Geographic Names (TGN) is a thesaurus, usually referenced within major, general-purpose, geographic databases and maps. TGN is intended to aid cataloging, research, and discovery of information about visual works and related topics. While most records in TGN include coordinates, these coordinates are approximate and are intended for reference only. The focus of TGN is on historical places, although enough information about the modern world is included to give context to historical places and to allow documentation and discovery of visual works.

>> https://www.getty.edu/research/tools/vocabularies/ tgn/ Iconclass: a classification system designed for art and iconography. It is the most widely accepted scientific tool for the description and retrieval of subjects represented in images (works of art, book illustrations, reproductions, photographs, etc.) and is used by museums and art institutions around the world. The Iconclass system is accessible through the Iconclass Browser and available as Linked Open Data (LOD).

>> http://www.iconclass.org/help/outline

♦ ULAN: the Getty Union List of Artist Names (ULAN) includes names, relationships, and biographical information for makers and other people and corporate bodies required for the documentation, collection, and discovery of information about art, architecture, and other material culture where the works are of the type collected by art museums and other repositories for visual cultural heritage, or that are architecture. Within scope are artists, architects, other makers, firms, and studios, both named and anonymous. Also included may be patrons and repositories of art.

> >> https://www.getty.edu/research/tools/vocabularies/ ulan/

perio.do: a public domain gazetteer of scholarly definitions of historical, art-historical, and archaeological periods. It eases the task of linking among datasets that define periods differently. It also helps scholars and students see where period definitions overlap or diverge.

>> https://perio.do/en/

d. Tools for Thesauri Development and Maintenance

Skosmos: a web-based tool providing services for accessing controlled vocabularies, which are used by indexers describing documents and searchers looking for suitable keywords. Vocabularies are accessed via SPARQL endpoints containing SKOS vocabularies.

>> https://skosmos.org/

Protégé: a free, open-source platform that provides a growing user community with a suite of tools to construct ontologies that represent domain models.

>> http://protege.stanford.edu/

TemaTres: an open source vocabulary server for managing controlled vocabularies, taxonomies and thesauri. It is an open source web tool, distributed under the GNU Public License (GPL) for the management and exploitation of controlled vocabularies, thesauri, taxonomies and other models of formal knowledge representation. Internally, it uses a model based on terms, which differs from a model based on concepts. On the other hand, it has features specially oriented at providing traceability data and quality control for the created models. It is an easy tool to use and is a possible starting point.

>> https://www.vocabularyserver.com/

Thmanager: an Open Source Tool for creating and visualizing SKOS RDF vocabularies, a W3C initiative for the representation of knowledge organization systems such as thesauri, classification schemes, subject heading lists, taxonomies, and other types of controlled vocabulary.

>> https://thmanager.sourceforge.io/

VocBench is an open source web platform (under the MPL Mozilla Public License) that allows for the collaborative editing of SKOS(XL) multilingual thesauri. This tool is mainly developed by the University of Rome Tor Vergata, and is closely related to the AGROVOC thesaurus.

>> http://vocbench.uniroma2.it/

e. Data Models and Interoperability Initiatives

CIDOC CRM (Conceptual Reference Model): built by ICOM's Documentation Committee, it is a theoretical and practical tool for information integration in the field of cultural heritage. It can help researchers, administrators and the public explore complex questions with regards to our past across diverse and dispersed datasets. The CIDOC CRM achieves this by providing definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation and of general interest for the guerving and exploration of such data. Such models are also known as formal ontologies. These formal descriptions allow the integration of data from multiple sources in a software and schema agnostic fashion.

>> http://www.cidoc-crm.org/

♦ LIDO (Lightweight Information Describing Objects), a format for contributing collection information for resource discovery within museums. The strength of LIDO lies in its ability to support the full range of descriptive information about museum objects. It can be used for all kinds of objects, e.g. art, architecture, cultural history, history of technology, and natural history. It is intended for delivering metadata, for use in a variety of online services, from an organization's online collections database to portals of aggregated resources, as well as exposing, sharing and connecting data on the web.

>> https://cidoc.mini.icom.museum/working-groups/lido/ what-is-lido/ Linked Art is a community working to create a shared model based on Linked Open Data to describe art. It also intends to implement that model in software and use it to provide content.

>> https://linked.art/

The **Europeana Data Model** is the basic structure for data intended for ingestion within Europeana.

>> https://pro.europeana.eu/page/edm-documentation

UK's Jisc supplies great advice on how your institution can take steps to improve interoperability between funder, publisher and institutional systems.

>> https://www.jisc.ac.uk/guides/supporting-open-access-through-metadata-and-improved-interoperability

A short overview about more data models commonly used in museums is available at

>> https://meta.wikimedia.org/wiki/GLAM/Metadata_ standards_and_Wikimedia#Overview_of_GLAM_metadata_standards_and_models

Data transfers to Wikimedia Commons are the ordinary procedure for sharing institutional archives in this global, open access repository. Basic guidelines for institutions interested in this process can be found at

>> https://meta.wikimedia.org/wiki/Grants:Learning_patterns/Data_transfers_to_Wikimedia_Commons:_Sharing_ institutional_archives

f. Other Tools and Technologies

Wikifier: a web service which takes a text document as input and annotates it with links to relevant Wikipedia concepts.

>> http://wikifier.org/

Ontome: Designed for any object-oriented structured data model, OntoME makes it easy to build, manage and align your ontology. It also aligns your classes to CIDOC CRM and compatible models and access to robust interoperability.

>> https://ontome.net/

IIIF, the International Image Interoperability Framework, is a set of open standards for delivering high-quality, attributed digital objects online at scale. Many of the images and audio/visual resources that are fundamental to research exist in silos, with access restricted to locally-built applications. IIIF provides freedom to work across those barriers. It is also an international community developing and implementing the IIIF APIs. It is backed by a consortium of leading cultural institutions.

>> https://iiif.io/

APIs (Application Programming Interfaces) are an alternative for content providers (like museums, and many others) willing to open access to digital files of the artifacts in their collections. Offering this kind of open data access encourages new visitor experience innovation, simplifies collaborations, offers machine-readability for digital art projects and provides an online outlet to showcase collections remotely, and more. Additional information and dozens of examples can be found at

>> https://www.programmableweb.com/category/museums

The Europeana APIs are a useful example. They enable developers to build applications that use the wealth of its collections drawn from the major museums and galleries across Europe. Their scope includes millions of cultural heritage items provided by more than 4,000 cultural institutions across Europe.

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

The Sustainability of Digital Formats (Library of Congress) website provides in-depth descriptions of over 440 formats sorted into content categories including: still image, sound, textual, moving image, Web archive, datasets, geospatial and generic formats with more to come. Important features of the format descriptions include documenting relationships between formats and factors to consider when evaluating formats including sustainability factors and quality and functionality factors.

>> https://www.loc.gov/preservation/digital/formats/

The **Recommended Formats Statement** from the Library of Congress identifies hierarchies of the physical and technical characteristics of creative formats, both analog and digital, which will best meet the needs of all concerned, maximizing the chances for survival and continued accessibility of creative content well into the future.

>> https://www.loc.gov/preservation/resources/rfs/



Our Top 5 Recommendations





Before digitizing your collection, consider the overall context. Examine your mission, your financial and human resources and how are you planning to use these data. This can facilitate decision-making at a more strategic level.



When digitizing, take into account the growing demand from various stakeholders (including educators, researchers, cultural and creative sector, IT) for rich metadata and high-quality digital items, allowing for their use and reuse within the context of both cultural heritage and other sectors. Standards are essential, so try to implement them in your institution as much as possible.



Think about long-term sustainability for your work. In-house solutions might be a way to begin, but they will not go far when expectations grow. Get serious advice for your software and format choices. Work to make your data FAIR-compliant, from the very beginning.



Facilitate access to and use of digitized cultural material as much as possible. Make sure to provide clear information on the user's rights, for instance by adopting Creative Commons licenses.



Facilitate access to and use of digitized cultural material as much as possible. Make sure to provide clear information on the user's rights, for instance by adopting Creative Commons licenses.

B Thank you!

We are very much indebted to all participants in the preparatory workshops, who provided ideas and support for the preparation of this document. It is our pleasure to thank these groups, made up of independent scholars as well as by professionals from the following museums:

- Museo Lázaro Galdiano, Madrid
- Museu de la Seda de Moncada
- Museo Nacional de Artes Decorativas, Valencia
- Museu del Disseny, Barcelona
- · Centro de Documentación y Museo Textil, Terrassa
- L'Etno, Valencia
- Patrimonio Nacional
- Conselleria d'Educació, Cultura i Esport (regional government, Valencia)
- Museo Diocesano di Caltanissetta
- Museo Diocesano di Cuneo
- Museo Diocesano di Monreale
- Museo Arcidiocesano "G.Boccanera" di Camerino
- Museo Diocesano di Genova
- Museo Basilica San Sebastiano
- Museo Diocesano di Massa Carrara Pontremoli
- Museo Diocesano di Padova
- Museo Diocesano di Reggio Calabria
- Museo Textil de Oaxaca, Mexico
- Philadelphia Museum of Art, United States
- Haus der Seidenkultur, Germany
- The Metropolitan Museum of Art, United States
- Klosterkammer Hannover, Germany
- Olympic Foundation for Culture and Heritage, Switzerland
- Fachhochschule Potsdam, Institute for Urban Futures, Germany
- State Silk Museum, Georgia
- Centre for Textile Research, University of Copenhagen, Denmark
- Mingei research project
- Museu Nacional do Traje, Portugal

SILKNOW. Silk heritage in the knowledge society: from punched cards to big data, deep learning and visual / tangible simulations.





Leibniz

Universität

EURECOM

Hannover

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 769504



"Jožef Stefan" Ljubljana, Slovenlja









2



For more information, please contact:

> Jorge Sebastian: jorge.sebastian@uv.es

> > Mar Gaitán: m.gaisal@uv.es