

A Workflow to Publish Collections as Data

Looking Back at Europeana.eu and Forward to the Common European Data Space for Cultural Heritage

Gustavo Candela^{1, 3, *}, Sally Chambers^{2, 3, 8}, Alba Irollo⁴, Nuno Freire⁴, Vicky Dritsou^{3, 5}, Antoine Isaac⁴, Agiatis Benardou^{3, 5}, Vicky Garnett^{3, 6}, Toma Tasovac^{3, 7}

¹ University of Alicante, Spain

² British Library, UK

³ Digital Research Infrastructure for the Arts and Humanities, DARIAH-EU

⁴ Europeana Foundation

⁵ Digital Curation Unit, IMSI/Athena RC, Greece

⁶ Trinity College Dublin, Ireland

⁷ Belgrade Center for Digital Humanities, Serbia

⁸ Ghent Centre for Digital Humanities, Ghent University, Belgium

* Corresponding author: Gustavo Candela; gcandela@ua.es

Transformations, A DARIAH Journal Abstract

Volume 1, 2025

<https://transformations.episciences.org>

Dates

Submitted: 15/11/2024

Accepted: 14/04/2025

Published: 25/06/2025

DOI:

[10.46298/transformations.14774](https://doi.org/10.46298/transformations.14774)

© The authors



Creative Commons Attribution 4.0
International

For decades, cultural heritage (CH) institutions have been making their digital collections available for potential communities of users. Recent advances in technology, such as machine learning, have provided a new context in which digital collections are a rich resource that can be analysed and reused by means of computational methods, for example in the humanities. Initiatives such as Collections as Data, the FAIR (findable, accessible, interoperable, reusable and CARE (collective benefit, authority to control, responsibility, ethics) data principles, and experimental Labs provide best practices and guidelines for publishing digital collections suitable for responsible computational use. In addition, data spaces have recently emerged as a new concept to foster creation, access and reuse of heritage data in which CH institutions play a key role as data providers. In this work we present a workflow for adopting Collections as Data in CH data spaces. The workflow has been developed in the context of the common European data space for cultural heritage and published on the Social Sciences and Humanities (SSH) Open Marketplace. It aims to support CH institutions, humanities researchers and computer scientists interested in making CH data available in data spaces. The article illustrates how the workflow can be adopted by CH institutions or applied by researchers wishing to publish digital collections suitable for computational use. We also demonstrate how the workflow is being adopted in the common European data space for cultural heritage, and describe a selection of potential areas that we believe will most benefit from the wide application of the workflow in the CH domain.

Keywords: digital libraries, Collections as Data, digital collections, digital humanities, workflows, metadata, data spaces, GLAM, cultural heritage data, datasets, research data

Μια Ροή Εργασιών για τη Δημοσίευση Συλλογών ως Δεδομένα: Αναδρομή στο Europeana.eu και Προοπτική προς τον Κοινό Ευρωπαϊκό Χώρο Δεδομένων για την Πολιτιστική Κληρονομιά

Περίληψη

Εδώ και αρκετές δεκαετίες, οι φορείς Πολιτιστικής Κληρονομιάς διαθέτουν τις ψηφιακές συλλογές τους διαδικτυακά, επιδιώκοντας τη διεύρυνση της προσβασιμότητάς τους και την αξιοποίησή τους από ένα ευρύ και διαφοροποιημένο σύνολο χρηστών. Οι πρόσφατες εξελίξεις στις τεχνολογίες της πληροφορίας, όπως η πρόοδος στον τομέα της Μηχανικής Μάθησης, έχουν διαμορφώσει ένα νέο τεχνολογικό πλαίσιο. Στο πλαίσιο αυτό, οι ψηφιακές συλλογές δεν αντιμετωπίζονται πλέον αποκλειστικά ως τεκμήρια πολιτιστικού περιεχομένου, αλλά αναγνωρίζονται ως πολύτιμα σύνολα δεδομένων, ικανά να αναλυθούν και να επαναχρησιμοποιηθούν μέσω υπολογιστικών μεθόδων στο πεδίο των Ψηφιακών Ανθρωπιστικών Επιστημών. Πρωτοβουλίες όπως η θεώρηση των Συλλογών ως Δεδομένα, οι αρχές FAIR (Findable, Accessible, Interoperable, Reusable) και CARE (Collective Benefit, Authority to Control, Responsibility, Ethics), καθώς και τα πειραματικά εργαστήρια, προσφέρουν κατευθυντήριες γραμμές και βέλτιστες πρακτικές για τη δημοσίευση ψηφιακών συλλογών με τρόπο που καθιστά εφικτή και αξιόπιστη την υπολογιστική αξιοποίησή τους. Στο ίδιο πλαίσιο, η έννοια των χώρων δεδομένων (data spaces) αναδύεται ως νέα βάση για την ενίσχυση της δημιουργίας, της πρόσβασης και της επανάχρησης δεδομένων Πολιτιστικής Κληρονομιάς, όπου οι φορείς της διαδραματίζουν κρίσιμο ρόλο ως πάροχοι δεδομένων. Στο παρόν άρθρο παρουσιάζουμε μια ροή εργασίας (workflow) για την εφαρμογή της προσέγγισης Συλλογές ως Δεδομένα (Collections as Data) εντός των χώρων δεδομένων Πολιτιστικής Κληρονομιάς. Η ροή αυτή έχει αναπτυχθεί στο πλαίσιο του Κοινού Ευρωπαϊκού Χώρου Δεδομένων για την Πολιτιστική Κληρονομιά και έχει δημοσιευθεί στην πλατφόρμα Social Sciences and Humanities Open Marketplace. Στόχος της είναι να υποστηρίξει φορείς Πολιτιστικής Κληρονομιάς, ερευνητές των Ανθρωπιστικών Επιστημών και επιστήμονες Πληροφορικής που επιδιώκουν τη συστηματική διάθεση και επαναχρησιμοποίηση δεδομένων Πολιτιστικής Κληρονομιάς μέσω υπολογιστικών διαδικασιών. Η παρούσα μελέτη εξετάζει τις δυνατότητες υιοθέτησης της εν λόγω ροής εργασίας τόσο από θεσμικούς φορείς όσο και από μεμονωμένους ερευνητές, και αποτυπώνει την εφαρμογή της στο πλαίσιο του Κοινού Ευρωπαϊκού Χώρου Δεδομένων για την Πολιτιστική Κληρονομιά. Επιπλέον, αναδεικνύει ενδεικτικούς τομείς εφαρμογής, οι οποίοι εκτιμάται ότι θα ωφεληθούν ουσιαστικά από την ευρύτερη υιοθέτηση της προτεινόμενης ροής εργασίας εντός του πεδίου της Πολιτιστικής Κληρονομιάς.

Λέξεις-κλειδιά: ψηφιακές βιβλιοθήκες, Συλλογές ως Δεδομένα, ψηφιακές συλλογές, Ψηφιακές Ανθρωπιστικές Επιστήμες, ροές εργασίας, μεταδεδομένα, χώροι δεδομένων, φορείς πολιτιστικής κληρονομιάς, δεδομένα πολιτιστικής κληρονομιάς, σύνολα δεδομένων, ερευνητικά δεδομένα

Acknowledgements

A reduced version of this work entitled “Validating a Reproducible Workflow for Publishing Collections as Data: The Case of Europeana” was presented at the DARIAH Annual Event 2024: Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities, Lisbon, NOVA FCSH, Universidade Nova de Lisboa, 19–21 June 2024. The initial version of the workflow provided in figure 1 was refined and enhanced as shown in figure 2 as a result of a workshop organised by Europeana and DARIAH within the data space consortium meeting, held in The Hague on 1–2 April 2025. We would like to thank the participants for their comments and suggestions.

Conflict of Interest

The authors do not have any conflict of interest to declare

INTRODUCTION

Data plays a crucial role in the digital transformation of society and institutions, serving as a key factor for the production of many goods and services. Its value depends on how it is created, shared and used (OECD 2022). Cultural heritage (CH) institutions host large, rich digital collections and for decades have made progress in making them publicly available. International CH initiatives have then brought these collections closer to the general public and professionals interested in reusing and aggregating data, and in increasing its findability and accessibility. Examples include the Digital Public Library of America,¹ the Australian platform Trove,² and, in the European context, the Europeana initiative.³ In the meantime, research infrastructures like DARIAH, the Digital Research Infrastructure for the Arts and Humanities,⁴ have emerged to support and enhance digitally enabled research. Advances in technology, especially in the field of machine learning, have paved the way to a new context in which CH data can be employed for computational use, such as computer vision (Candela et al. 2025), which is proving to be ground-breaking in the humanities.

Data spaces provide new platforms to facilitate and foster the exchange of data collections between organisations. A data space is defined as a “decentralised infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles” (Brunet et al. 2022). In this panorama, data spaces are key to accelerating the digital transformation in society, fostering the publication and reuse of data (Franklin, Halevy, and Maier 2005). They aim to provide data-sharing mechanisms and means to ensure providers remain in control of their data. In 2020, the European Commission proposed a data strategy to facilitate the creation of data spaces across the EU in relevant sectors such as health, the Green Deal and public administration, as well as a data space for cultural heritage (European Commission 2020). Some initiatives have been ahead of their time, especially in research and innovation. One such example is the European Open Science Cloud (EOSC),⁵ launched in 2018 as an infrastructure that enables researchers to store, share, process, analyse and reuse data within and across disciplines and borders. In the near future, the European Collaborative Cloud for Cultural Heritage (ECCCH), implemented through the project European Cloud for Heritage Open Science (ECHOES),⁶ will connect researchers and professionals from the CH sector by providing integrated services and tools using cutting-edge technologies.

Several data space initiatives have explored the use of CH content (Dobrev, Stefanov, and Ivanova 2022; Brunet et al. 2022; Niccolucci, Felicetti, and Hermon 2022). In the context of the common European data space for cultural heritage, DARIAH, in partnership with the Europeana Foundation (hereafter, Europeana), has developed and published a workflow to support the publication and reuse of CH data suitable for computational use (Candela, Chambers, and Irollo 2023).

-
1. <https://dp.la>
 2. <https://trove.nla.gov.au>
 3. <https://www.europeana.eu>
 4. <https://www.dariah.eu>
 5. <https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud>
 6. <https://www.echoes-eccch.eu/>

The objective of this partnership is to support the needs of researchers and CH professionals who are interested in reusing data for academic research, teaching and learning. Additionally, it raises awareness of these needs among data providers and informs the data supply process. Development of the workflow has been guided by the following questions: What are the steps required to create datasets suitable for computational use? What are the challenges? To what extent could we contribute to the establishment of best practices and guidelines in the data space context?

Hosted on the Social Sciences and Humanities (SSH) Open Marketplace,⁷ the workflow fully embraces the open science principles promoted by that platform, proposing itself as an infinitely expandable structure thanks to the contribution of the community of users. Published in its first version in August 2023, the workflow builds on previous work carried out within the International GLAM Labs Community, a community of professionals who believe in the potential of Labs to foster advanced data reuse in the context of GLAMs (galleries, libraries, archives and museums).⁸ Like many Labs experiments, the workflow takes inspiration from—and contributes to—the Collections as Data movement, taking its steps from a project of the same name that ran from 2018 to 2023 and resulted in a series of principles (Padilla et al. 2019b; Padilla et al. 2023). In general, the content provided by CH institutions is made available in the form of a dataset (Ames 2020). The Collections as Data movement promotes the publication of digital collections that support computationally driven research and teaching (Padilla et al. 2019a; Mahey et al. 2019). From this perspective, as the workflow suggests, a dataset consists of a structured collection of data facilitating computational access to address specific research questions and offering content in different formats (Candela et al. 2024).

This work intends to answer the following questions: (i) how can we support institutions in applying the workflow by providing examples of application, (ii) what are the benefits for institutions in following the steps proposed by the workflow, and (iii) how can the workflow be refined and enhanced based on practical experience within Europeana.

While a pilot based on the workflow is in the pipeline to publish new types of data in the common European data space for cultural heritage, this article aims to present the experiments and assessments conducted since publication of the workflow, providing an analysis of the potential benefits of its wider application. Since the deployment of the data space builds on the 15-year experience developed through publishing CH data on Europeana.eu, particular attention is paid to the efforts made by Europeana to publish data that can be processed and analysed by machines. These efforts are presented here and discussed against the series of steps suggested by the workflow, although they matured independently and, in more recent years, have been aligned with the FAIR (findable, accessible, interoperable, reusable) Principles fostered in the context of the EOSC.

The paper is organised as follows: after a brief description of the state of the art, section describes the Workflow to Publish Collections as Data. Section discusses some characteristics of Europeana.eu against the steps suggested by the work-

7. <https://marketplace.sshopencloud.eu/workflow/I3JvP6>

8. <https://glamlabs.io/>

flow. Finally, section provides an analysis of the benefits for a wide diversity of contexts in which the workflow could be a useful and valuable resource.

BACKGROUND

This section presents previous approaches relating to the provision of digital collections suitable for computational use, as well as the Workflow to Publish Collections as Data in the context of data spaces.

PUBLISHING COLLECTIONS AS DATA IN CH INSTITUTIONS

CH institutions have been exploring new ways to make their digital collections available for computational use by means of a wide variety of approaches. Besides the Collections as Data principles specifically conceived to be implemented in CH institutions (Chambers et al. 2023), the FAIR Principles (Wilkinson et al. 2016), conceived primarily for research data, have been a point of reference for CH institutions, encouraging them to improve the findability, accessibility, interoperability and reusability of digital collections. While the FAIR Principles focus on technical and scientific aspects of data stewardship, the CARE (collective benefit, authority to control, responsibility, ethics) Principles complement this perspective by aiming to serve indigenous data governance initiatives (Carroll et al. 2020). Several projects are based on the extraction, packaging and reuse of datasets (International GLAM Labs Community 2022; Sherratt 2021). Experimental Labs have been implemented at GLAMs to foster and showcase innovative examples of reuse of their digital collections, including a wide diversity of content in the form of images and text (Mahey et al. 2019). For example, the Data Foundry at the National Library of Scotland publishes data openly and in reusable formats (Ames 2020); KBR, the Royal Library of Belgium, has worked towards facilitating data-level access to digitised and born-digital collections for digital humanities research (Royal Library of Belgium 2020); and the Library of Congress has invested in the creation of crowdsourcing datasets (Hynning et al. 2022). Specific projects have focused on publishing newspapers as data⁹ or decolonising CH content (Candela, Pereda, et al. 2023). There is increasing interest in making documentation about digital collections available in the form of datasheets (Alkemade, Claeysens, Colavizza, Freire, Lehmann, et al. 2023), considering aspects such as provenance, data quality and examples of use, and taking into account best practices matured in other sectors.

In combination with these approaches, metadata vocabularies can be used to describe and enrich datasets with external repositories (Ducatteeuw et al. 2023). Such vocabularies are based on the use of the Semantic Web and Linked Data principles (Berners-Lee et al. 2000). For example, the vocabularies Schema.org and Data Catalog Vocabulary (DCAT) provide classes and properties to describe datasets, as well as enabling interoperability between them (World Wide Web Consortium 2024). Other vocabularies and conceptual models, such as the Europeana Data Model (EDM) (Doerr et al. 2010) and the CIDOC Conceptual Reference Model (CRM) ontology (Bekiari et al. 2021), address the description of metadata. Furthermore, a data model for the European Collaborative Cloud for Cultural Heritage is currently under development.

9. <https://dlocasdata.domains.uflib.ufl.edu/>

In this context, the need for workflows that facilitate the adoption of best practices has made inroads in the CH sector. To develop a workflow, it is fundamental to resort to an online interface (Puren et al. 2023) that allows users to include a description of the workflow, associate metadata, create and describe the steps of the workflow, and reference relevant resources. Workflows enable users to gain a better understanding of complex processes by breaking them down into structured and individual tasks, while also systemising their execution, like, for example, a workflow proposed in the field of computational historical linguistics (Wu et al. 2020). Among the platforms that have emerged to create workflows, the SSH Open Marketplace,¹⁰ built as part of the Social Sciences and Humanities Open Cloud project (SSHOC), also enables the publication of tools, services, training materials and datasets (Barbot et al. 2022; Barbot et al. 2024). This platform potentially adds value to the Collections as Data workflow presented here, as the variety of resources published on the platform, by their very nature, can be of interest to the CH sector and therefore selected and used to enrich the workflow itself, making it ultimately the result of a collaborative effort.

THE WORKFLOW TO PUBLISH COLLECTIONS AS DATA

Several initiatives have recently emerged to promote data exchange and reuse in the context of data spaces. For example, the International Data Spaces Association (IDSA)¹¹ focuses on defining a global standard for international data spaces (IDS) and fostering the related technologies that will drive the data economy of the future across industries. In this context, the IDS Reference Architecture Model (IDS-RAM) establishes a five-layer structure (business, functional, process, information and system) addressed from the perspective of security, certification and governance (International Data Spaces Association 2019). In particular, the information layer addresses the use of shared vocabularies and data schema to describe the content, while the system layer describes the data service, which offers an Application Programming Interface (API) to store, access or process data.

Deployed by a consortium established in 2022 under the stewardship of European, the common European data space for cultural heritage has opened up new perspectives regarding publishing, sharing and reusing CH data, with a strong focus on decentralisation (in terms of governance) and datasets (in terms of content). The Workflow to Publish Collections as Data was conceived for datasets suitable for computational use, taking into account the leading concepts on which data spaces are built. As shown in figure 1, the steps suggested by the workflow reflect specific tasks and are grouped according to the information and system layers defined in the IDS-RAM (International Data Spaces Association 2019).¹² It represents a further and relevant development of a checklist to publish Collections as Data, put together and tested within the International GLAM Labs Community in 2022 as an easy-to-follow approach to support small and medium organisations in publishing their digital collections (Candela, Gabriëls, et al. 2023). Figure 2 shows a refinement of the original workflow based on comments and suggestions gath-

10. <https://marketplace.sshopencloud.eu/>

11. <https://internationaldataspaces.org>

12. The original checklist was composed of 10 items; the GLAM Labs Community suggested including an additional one concerning the provision of terms of use, expanding the checklist to 11. As terms of use could be grouped within an existing item (namely, “licence”), the workflow once again includes 10 steps.

ered from the consortium behind the common European data space for cultural heritage in April 2025. Notably, the steps have been grouped according to specific “stages” aimed at further clarifying how the workflow can be applied. By assessing, redefining and enriching the steps suggested by the checklist, the workflow presents the following key aspects:¹³

- In general, it considers the digital readiness of organisations as well as the available resources in terms of staff and technology, since it provides a set of 10 steps that can be applied in a flexible way. Note that there are no mandatory items or strict order when applying the workflow steps. Figure 2 shows a “loop” that suggests revisiting the work done in earlier steps in light of later ones, if needed. These aspects can be considered according to the needs and requirements of the institution. For example, an institution with limited IT architecture and resources can make the data available in the form of ZIP files using its website or a public platform such as GitHub. In other cases, focusing on particular groups of steps at once might be interesting since they may be related (e.g. examples of use, documentation and licences).
- It promotes the adoption of permissive licences (e.g. Creative Commons) and terms of use to facilitate reuse of digital collections by the community of users and encourage responsible data use. However, in some cases, several licences might be needed because different objects in the dataset, as well as the metadata describing it, may be covered by separate agreements.
- It fosters the inclusion of a suggested citation format to help users properly cite the dataset and give credit, recognition and visibility to the institution. A simple and usable way of providing a citation is in the form of text that can be used in the bibliography of an article. The workflow also promotes the use of persistent identifiers, such as a DOI, Archival Resource Key (ARK) or Handle, to link to the dataset.
- It addresses the information and (to a certain extent) system layers proposed by the IDS-RAM. Additional work is required to address the business, functional and process layers proposed by that model. Data spaces are based on the use of connectors to provide access to data in a federated environment. This functionality can be fulfilled, to some extent, by the APIs used by institutions (see the step concerning the provision of an API in figure 1).
- It considers an easy-to-follow data-sharing life cycle based on existing public platforms (e.g. GitHub and Zenodo) that can be used to publish and make the datasets available as recommended in the related step. Institutions can decide how they want to make their digital collections available in terms of their technical infrastructure and resources.
- It fosters the development and improvement of digital skills by suggesting the inclusion of examples of use based on reproducible code, such as Jupyter Notebooks and documentation. Additional information concerning data quality, provenance and descriptions of how the digital collection was curated and published can be provided in the form of datasheets (Alkemade, Claeysens, Colavizza, Freire, Lehmann, et al. 2023). However, other approaches can be used to provide this type of information, for example by publishing a portal page dedicated to the dataset. This contextual information is crucial for researchers interested in reusing the datasets.

13. Additional information and examples about each step of the workflow are provided in the SSH Open Marketplace (Candela, Chambers, and Irollo 2023).

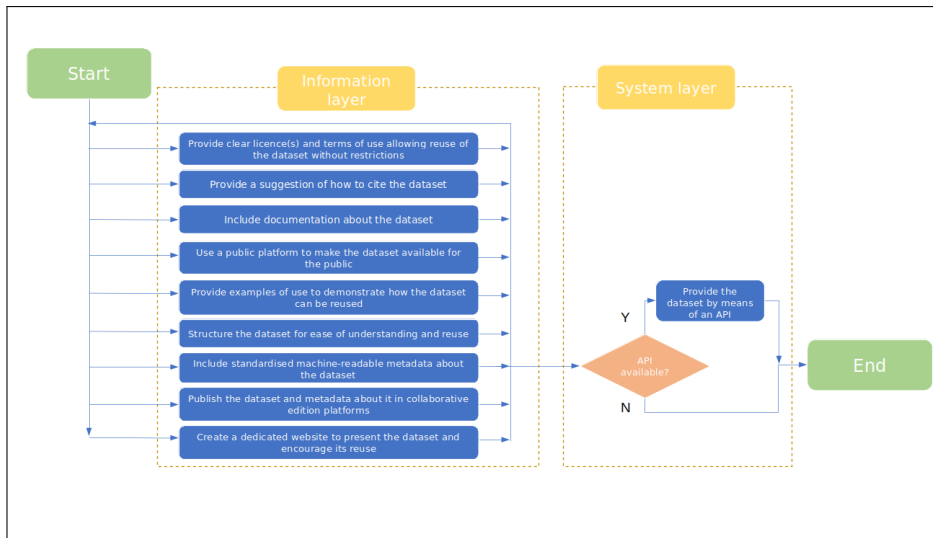


Figure 1: Description of the steps provided by the Workflow to Publish Collections as Data, available at the SSH Open Marketplace (Candela, Chambers, and Irollo 2023).

- Concerning the overall structure, it provides guidelines, examples and initiatives made available by relevant institutions to ensure a coherent internal organisation of the dataset.
- The step on standardised and machine-readable metadata provides guidelines to foster interoperability by means of controlled vocabularies (e.g. DCAT). Note that the metadata can include details about the content and provenance of the dataset.
- It promotes the use of collaborative platforms such as Wikidata and the SSH Open Marketplace to describe and increase the visibility of the dataset.

TOWARDS IMPLEMENTATION OF THE WORKFLOW ON EUROPEANA.EU

The common European data space for cultural heritage is built on the foundation of Europeana.eu, which provides access to over 60 million digitised items shared by approximately 4,000 CH institutions across Europe—from texts and images to audiovisual material and 3D items. Professionals, students and citizens can explore and reuse this data in innovative ways. Researchers and teachers are among the largest user groups of Europeana.eu.

Europeana’s commitment to facilitating the reuse of data is long lasting and has been realised, for example, through its work on rights statements. The metadata received by Europeana from the providing institutions always contains rights-related information for each of the items published. To ensure CH data is reused in accordance with rights and make it simpler to understand what is allowed, Europeana adopted specific rights statements articulated in the Europeana Licensing Framework¹⁴ (i.e. Creative Commons licences, Creative Commons public domain tools, and rights statements by the RightsStatements.org consortium¹⁵).

14. <https://pro.europeana.eu/page/europeana-licensing-framework>

15. <https://rightsstatements.org/>

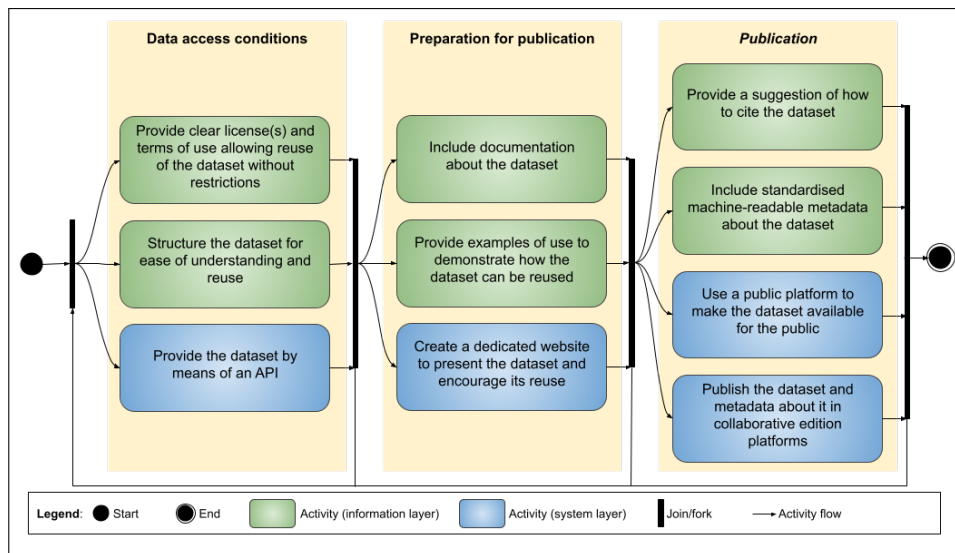


Figure 2: Refinement of the original workflow provided in figure 1. The steps have been grouped according to particular tasks in order to further clarify how the workflow can be applied.

Concomitantly with the launch of the EOSC in 2018, Europeana has taken increasing and consistent measures to ensure CH data complies with the FAIR Principles. A significant step forward in this direction has been to explore the possibilities offered by adopting persistent identifiers (PIDs) at a dataset level. Relying as much as possible on unique and persistent identifiers is a key aspect for building a space where data can be easily shared for reuse. This factor motivated Europeana to also investigate the use of PIDs, analysing how they are used to identify CH objects in the metadata received by CH institutions (Freire et al. 2023). They found that PIDs are currently used in 13 percent of the records. Europeana also issued a policy proposal for PIDs¹⁶ and initiated an assessment with various stakeholders. The policy establishes clear expectations for data partners seeking to implement and contribute PIDs. As a further step, PIDs must be integrated in the citations automatically suggested by the Europeana.eu website when users download images (a measure also in line with one of the steps in the Workflow to Publish Collections as Data).

Europeana also participates in the International GLAM Labs Community. In 2022, it helped define and test the original Collections as Data checklist designed for small and medium-sized institutions. For the present article, Europeana has assessed to what extent Europeana.eu already complies with the steps suggested by the workflow. This complements the input that Europeana collects from stakeholders in academia and research. This input brings new ideas for research and development, engineering, data publishing services, and user experience. Thanks to the assessment, we have identified that Europeana.eu already complies—or is close to complying—with the following areas through functionality provided by its existing systems, driven by other motivations than the workflow itself, but which turn out to be well aligned with it:

16. <https://pro.europeana.eu/index.php/post/policy-for-persistent-identifiers-in-the-data-space>

- Supporting dataset documentation—To improve the practice of documenting CH datasets, a working group¹⁷ was set up within the Europeana Research¹⁸ and EuropeanaTech¹⁹ communities, themselves part of the Europeana Network Association. The working group built upon the concept of datasheets, which has gained traction in the machine learning community (Geburu et al. 2021). Datasheets provide context and information about their content in order to avoid some of the concerns arising with their use in machine learning. Since CH datasets are characteristically different from datasets in other sectors, the working group defined a specific datasheet template for them (Alkemade, Claeysens, Colavizza, Freire, Irollo, et al. 2023; Alkemade, Claeysens, Colavizza, Freire, Lehmann, et al. 2023). It is intended to guide dataset creators in providing information about the collection process, recommended uses and potential societal biases in the data, and will hopefully help to avoid the concerns identified in the past regarding inappropriate use of datasets. The template has already been put into practice (Lehmann and Schneider 2024).
- Supporting the creation of machine-readable metadata for the datasets—In recent years, Europeana has explored the application of vocabularies to provide machine-readable metadata for CH datasets (Freire 2020), namely with Schema.org and the Data Catalog Vocabulary (DCAT) (World Wide Web Consortium 2024). More recently still, Europeana has worked with the DCAT Application Profile for data portals in Europe (DCAT-AP), used to describe public sector datasets in Europe. This work resulted in the publication of the Europeana.eu dataset on data.europa.eu (Europeana Foundation 2008), the official portal for European data. The application of DCAT for the Workflow to Publish Collections as Data is still in the requirements-analysis stage. However, Europeana’s work with DCAT-AP is expected to provide a solid basis for providing metadata for CH datasets. The mapping was designed to be as comprehensive as possible given the information available on the datasets. The mapping covers 13 classes used in DCAT-AP, adopts most of the vocabularies recommended in DCAT-AP and uses machine-readable rights statements. More details about the mapping are provided in ???. The validity of the DCAT-AP metadata created by Europeana was checked using the validator service²⁰ provided by data.europa.eu, and its data quality measured on the Metadata Quality Assurance (MQA) dashboard, also provided by data.europa.eu. High scores were achieved for most of the quality ratings available in the MQA dashboard, providing robust guarantees that the dataset metadata is machine-actionable and informative. The screenshot in figure 3 illustrates the Europeana.eu dataset on the MQA dashboard and shows the scores obtained for the quality ratings employed there.
- Using a public platform to make datasets available to the public—The publication of CH datasets should be done on specialised public platforms that support versioning and persistent identification (for example with a DOI). Re-

17. The Datasheets for Digital Cultural Heritage Working Group: <https://pro.europeana.eu/project/datasheets-for-digital-cultural-heritage-working-group>.

18. <https://pro.europeana.eu/page/europeana-research>

19. <https://pro.europeana.eu/page/europeanatech>

20. <https://data.europa.eu/mqa/shacl-validator-ui/>

cently, several platforms have emerged. Some examples are Zenodo,²¹ DataCite²² and Hugging Face.²³ Europeana has already published a number of datasets via Zenodo,²⁴ while some Europeana-related data can be found on Hugging Face. However, these are either the result of “satellite activities” or may not necessarily be about the data on Europeana.eu. In addition, a platform specialised in CH—AI4Culture²⁵—has recently been established, providing access to a pool of curated AI software tools, datasets and capacity-building material for CH. Europeana expects to connect more systematically to one or more of these repositories for publishing curated datasets.

- Data access via APIs—Europeana.eu already provides several (standard) APIs²⁶ that facilitate reuse of its data. These make the item-level data immediately available. This data can potentially be used for datasets built according to the Workflow to Publish Collections as Data. A key API for the implementation of the workflow is the User Set API.²⁷ Originally designed to provide access to the sets of objects that end-users bookmark via the Europeana.eu website, or which Europeana wants to feature in editorial content, it is currently being investigated to support datasets that align with the Collections as Data approach, engaging CH institutions and users in a different level of data curation, and offering users access and possibility of reuse at dataset level beyond the Europeana APIs.

A specific area of analysis and experimentation concerns the curation of new “user sets” from the last item. Although the workflow was originally defined with CH institutions in mind, a large aggregation of CH data like Europeana.eu makes it possible to apply the workflow in a new scenario, where new cross-institutional datasets are built by professionals, such as researchers who locate and collect data from multiple CH institutions according to their topic of interest via the Europeana website.²⁸ These new datasets have great potential for reuse, despite being less official, generally smaller and very diverse.

DISCUSSION

The workflow discussed in section can be used in a wide diversity of contexts. From this perspective, and in order to foster its use and adoption, we see several areas in which the workflow could be a useful and valuable resource. Nevertheless, due to the wide diversity of content made available by institutions, there is still room for customisation and improvement. This section starts by presenting the key areas positively impacted by the workflow and follows with a discussion of its limitations.

21. <https://zenodo.org/>

22. <https://datacite.org/>

23. <https://huggingface.co/>

24. https://zenodo.org/communities/europeana/records?q=&f=resource_type%3Adataset

25. <https://ai4culture.eu/>

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

27. <https://europeana.atlassian.net/wiki/spaces/EF/pages/2486960140/User+Set+API+Documentation>

28. These datasets can be hand-picked in “user galleries” or built with the help of queries against the Europeana.eu database (for example, objects created by a particular author).

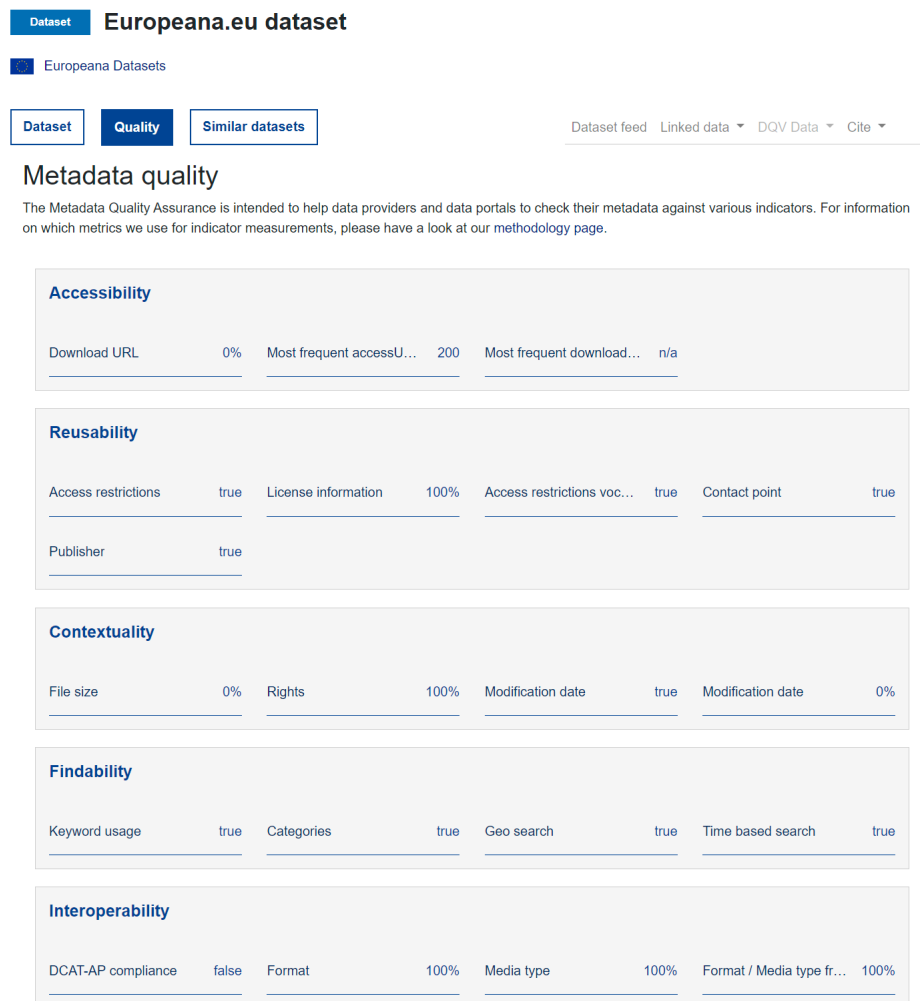


Figure 3: Results for the description of Europeana.eu's dataset from the data.europa.eu Metadata Quality Assurance (MQA) dashboard: <https://data.europa.eu/data/datasets/http-data-europeana-eu-dataset-europeana/quality>.

BENEFITS OF THE WORKFLOW FOR DATA REUSE

The workflow can help establish best practices and guidelines in terms of digital collections and their reuse in data-space-related platforms. It can also strengthen and expand the community of data reusers. In this section, we present the key areas where we foresee significant benefits from applying the workflow. Note that most of these areas have already been identified and recently explored by Europeana, and that the workflow presents an additional way for Europeana to further contribute to them.

Access to CH data. While tools and prototypes in GLAM institutions facilitate the use of data without having to parse and process the content, researchers in some cases prefer retrieving and processing the content using their own methods, techniques and configuration parameters. Data spaces are platforms in which federated data can be made available, enabling interoperability, data sharing, integration and reuse while ensuring security and trustworthiness. The workflow can be used to identify best practices for making data available, particularly for small and medium-sized institutions, also by taking into account recently introduced tools and services, such as Eclipse Dataspace Components,²⁹ which facilitate communication between institutions and data spaces. This is a key benefit for the common European data space for cultural heritage, given its mission to disseminate and facilitate the reuse of CH data.

Higher Education. Computer science, multimedia and digital humanities courses can be improved by including data spaces as data sources to enable several use cases, and by including the proposed workflow as additional educational material. Several tasks could be performed in this context, such as describing how the workflow could be used in particular scenarios, adapting it to other fields and further developing each of the tasks within the workflow. Initiatives like DARIAH-Campus³⁰ provide a discovery framework and hosting platform for learning resources. As part of the deployment of the common European data space for cultural heritage, it has been used to cover a range of key topics relating to digital CH relevant to higher education and research. The first courses focus on CH data, CH data modelling (with a particular focus on the EDM) and Europeana APIs (Dritsou et al. 2024a, 2024b, 2024c). A specific course covers Collections as Data (Candela et al. 2024).

Training materials for capacity building. The proposed workflow is valuable as training material, as it offers structured guidelines for engaging effectively with data spaces. Developing it as training material is essential to foster the use of data spaces³¹ and promote the adoption of the workflow by CH institutions. The aim could be to create a comprehensive framework for CH professionals, as well as for students aiming to pursue a career in the sector and preparing for the opportunities and challenges raised by digital cultural heritage.³² In 2024, Europeana and DARIAH organised a training workshop that attracted much interest among

29. <https://projects.eclipse.org/projects/technology.edc>

30. <https://campus.dariah.eu>

31. <https://pro.europeana.eu/event/collections-as-data-collaborating-across-data-spaces-for-cultural-heritage-and-open-science>

32. <https://pro.europeana.eu/post/exploring-digital-cultural-heritage-as-an-emerging-university-subject>

CH professionals.³³ It was followed by an event on Collections as Data at KBR, the Royal Library of Belgium, tailored to national library professionals. Workshops at relevant conferences can help not only to boost the use of data spaces in CH institutions, but also to enrich existing workflows and materials by identifying additional needs and requirements (Digital Research Infrastructure for the Arts and Humanities and Ferguson 2024). For example, in a workshop recently held at the International Conference on Theory and Practice of Digital Libraries (TPDL 2024), the workflow discussed in this article was used as a case study.³⁴

Data cloud environments for CH. A virtual research environment that also offers computing resources for running code (e.g. Jupyter Notebooks) and supports collaborative work among researchers and practitioners requires rich data (Brunet et al. 2022). The workflow could be used as a set of guidelines to provide the data in a cloud environment, fostering its reuse in innovative and creative ways. In addition, the use of cloud environments facilitates the integration of data sources, avoiding the creation of data silos and facilitating the interconnection of different repositories to create a unified view.

SCOPE, KNOWN LIMITATIONS AND ONGOING RESEARCH

This work provides a starting point for adopting Collections as Data in CH data spaces. However, and due to the wide diversity of content made available by institutions, there is still room for customisation and improvement. Each of the steps provided within the workflow can be adapted to meet the particular requirements of an organisation. In cases where the workflow is utilised by large aggregators of CH data, like Europeana.eu, adaptation is essential, as each aggregation network operates under specific terms, conditions and data infrastructures. This work focuses on the items included in the checklist, organised according to the information and system layers provided by the IDS-RAM. To fully exploit each of these layers (e.g. ingestion, access control and sustainability), additional elements should be considered. Applying the workflow in data spaces requires analysis of the content provided by datasets to ensure efficient processing for all interested parties.

Assessing the impact of digital collections in society and academia is crucial if we are to better understand how they are reused and identify potential new needs and requirements from the community (Shiri et al. 2023). In this sense, further work is required in order to explore the inclusion of additional steps in the workflow to integrate existing frameworks for assessing digital content reuse, such as D-CRAFT.³⁵

Automating the workflow could facilitate its adoption by CH institutions. While automating the entire process is a complex and difficult task, some steps could be supported using different means. For instance, a web interface could facilitate the creation of machine-readable metadata according to specific initiatives (Alkemada, Claeysens, Colavizza, Freire, Lehmann, et al. 2023). Other examples could draw on automatic data quality analysis to provide additional documentation. In

33. <https://pro.europeana.eu/index.php/event/a-workflow-to-publish-collections-as-data-training-workshop>

34. IMPACT workshop held at TPDL 2024, available at <https://tpdl2024.nuk.si/index.html#content1-e9>.

35. <https://reuse.diglib.org/>

addition, the generation of code to create prototypes and examples of use based on the content could also be provided.

Adoption of the workflow could be facilitated if it is specified in a standard machine-readable way. The SSH Open Marketplace enables the publication of resources such as workflows using a web interface; however, the machine-readable information provided is limited to the metadata describing the resource. Although the SSH Open Marketplace was used to publish the workflow due to its potential in the humanities, additional work is required to express (future versions of) the workflow in a machine-readable way. For example, the development of “composable” research workflows based on open science methods is being explored in the OSCARS project.³⁶

CONCLUSIONS

Over the past few years there has been growing interest in the rich content provided by CH institutions, particularly those in the GLAM sector. As part of their preservation and curation work, they have explored new ways to make their digital collections suitable for computational use following the Collections as Data, FAIR and CARE principles. Data spaces have emerged as new platforms to include interoperable datasets to foster data sharing and reuse across disciplines. The Collections as Data principles and workflow reflecting them are now considered a set of guidelines and best practices in the common European data space for cultural heritage, and will inform its deployment in the coming years.

Future work will focus on automating and improving the workflow by including additional features, such as a carbon footprint assessment, and by addressing ethical issues and quality. We will also align this work with European initiatives such as the EOSC, ECHOES and OSCARS. In addition, further user feedback will be collected and evaluation performed in order to improve the workflow. Finally, the use of an ontology, such as the Publishing Workflow Ontology (PWO) (Gangemi et al. 2014) and the Business Process Model (BPM) (Russell et al. 2006), will be explored to provide a machine-readable version of the workflow.

36. <https://oscars-project.eu/oscars-work-packages>

REFERENCES

- Alkemade, Henk, Steven Claeysens, Giovanni Colavizza, Nuno Freire, Alba Irollo, Jörg Lehmann, Clemens Neudecker, Giulia Osti, and Daniel van Strien. 2023. "Datasheets for Digital Cultural Heritage Datasets." Zenodo. Version 1, September. <https://doi.org/10.5281/zenodo.8375034>.
- Alkemade, Henk, Steven Claeysens, Giovanni Colavizza, Nuno Freire, Jörg Lehmann, Clemens Neudecker, Giulia Osti, and Daniel van Strien. 2023. "Datasheets for Digital Cultural Heritage Datasets." *Journal of Open Humanities Data* 9 (1). <https://doi.org/10.5334/johd.124>.
- Ames, Sarah. 2020. "Developing the Data Foundry: The National Library of Scotland's Data-delivery Platform." In *15th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2020 (Ottawa, Canada, 20–25 2020), Conference Abstracts*, edited by Laura Estill and Jennifer Guiliano. <https://dh-abstracts.library.virginia.edu/works/9251>.
- Barbot, Laure, Maja Dolinar, Edward J. Gray, Cristina Grisot, Klaus Illmayer, Michael Kurzmeier, and Barbara McGillivray. 2024. "Contextualizing Research Tools & Services Through Workflows in the SSH Open Marketplace." *Journal of Open Humanities Data* 10 (1). <https://doi.org/10.5334/johd.192>.
- Barbot, Laure, Edward Gray, Frank Fischer, Matej Ďurčo, Alexander König, Marie Puren, Stefan Buddenbohm, Cesare Concordia, and Klaus Illmayer. 2022. "The SSH Open Marketplace: A Multi-voiced Story." Poster presented at Storytelling—DARIAH Annual Event 2022 (Athens, Greece and online, 31 May to 3 June). Zenodo. Version 1, May. <https://doi.org/10.5281/zenodo.6580303>.
- Bekiari, Chrysoula, George Bruseker, Martin Doerr, Christian-Emil Ore, Stephen Stead, and Athanasios Velios. 2021. *Definition of the CIDOC Conceptual Reference Model*. Version v7.1.1. CIDOC CRM Special Interest Group. <https://doi.org/10.26225/FDZH-X261>.
- Berners-Lee, Tim, Dan Connolly, Lynn Andrea Stein, and Ralph Swick. 2000. "The Semantic Web." W3C. Updated 15 August. <https://www.w3.org/2000/Talks/0906-xmlweb-tbl/text.htm>.
- Brunet, Pere, Livio De Luca, Eero Hyvönen, Adeline Joffres, Peter Plassmeyer, Martin Pronk, Roberto Scopigno, and Gábor Sonkoly. 2022. *Report on a European Collaborative Cloud for Cultural Heritage: Ex-Ante Impact Assessment*. Directorate-General for Research and Innovation (European Commission). <https://doi.org/10.2777/64014>.
- Candela, Gustavo, Sally Chambers, and Alba Irollo. 2023. "A Workflow to Publish Collections as Data: The Case of Cultural Heritage Data Spaces." SSH Open Marketplace. <https://marketplace.sshopencloud.eu/workflow/I3JvP6>.
- Candela, Gustavo, Vicky Dritsou, Sally Chambers, Agiatis Benardou, Alba Irollo, and Toma Tasovac. 2024. "Introduction to Collections as Data." DARIAH-Campus Training Module. Version 1.0.0. <https://hdl.handle.net/21.11159/019595c5-1bc8-751c-8201-bcdb34955a96>.
- Candela, Gustavo, Nele Gabriëls, Sally Chambers, Milena Dobrev, Sarah Ames, Meghan Ferriter, Neil Fitzgerald, et al. 2023. "A Checklist to Publish Collections as Data in GLAM Institutions." *Global Knowledge, Memory and Communication* (November). <https://doi.org/10.1108/GKMC-06-2023-0195>.
- Candela, Gustavo, Olga Holownia, Max Odsbjerg, Mirjam Cuper, Nele Gabriëls, Katrine Hofmann, Edward J. Gray, Sally Chambers, and Mahendra Mahey. 2025. "Promoting Computational Access to Digital Collections in the Nordic and Baltic Countries: An Icelandic Use Case." *Journal of Open Humanities Data* 11 (1). <https://doi.org/10.5334/johd.261>.
- Candela, Gustavo, Javier Pereda, Dolores Sáez, Pilar Escobar, Alexander Sánchez, Andrés Villa Torres, Albert A. Palacios, Kelly McDonough, and Patricia Murrieta-Flores. 2023. "An Ontological Approach for Unlocking the Colonial Archive." *ACM Journal on Computing and Cultural Heritage* 16 (4): 1–18. <https://doi.org/10.1145/3594727>.

- Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark A. Parsons, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19 (1): 43. <https://doi.org/10.5334/dsj-2020-043>.
- Chambers, Sally, Melanie Walsh, Michelle Caswell, Geoff Harder, Mercedes Okumura, Julia Corrin, Gabriela Baeza Ventura, et al. 2023. "Position Statements—Collections as Data: State of the Field and Future Directions." Zenodo. Version 1, May. <https://doi.org/10.5281/zenodo.7897735>.
- Digital Research Infrastructure for the Arts and Humanities and Kim Ferguson. 2024. "Book of Abstracts, DARIAH AE 2024—Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities." Zenodo. Version v2, September. <https://doi.org/10.5281/zenodo.13655424>.
- Dobreva, Milena, Krassen Stefanov, and Krasimira Ivanova. 2022. "Data Spaces for Cultural Heritage: Insights from GLAM Innovation Labs." In *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries—24th International Conference on Asian Digital Libraries, ICADL 2022 (Hanoi, Vietnam, 30 November–2 December 2022)*, *Proceedings*, 492–500. Lecture Notes in Computer Science. Cham: Springer. https://doi.org/10.1007/978-3-031-21756-2_41.
- Doerr, Martin, Stefan Gradmann, Steffen Henricke, Antoine Isaac, Carlo Meghini, and Herbert van de Sompel. 2010. "The Europeana Data Model (EDM)." In *World Library and Information Congress: 76th IFLA General Conference and Assembly (Gothenburg, Sweden, 10–15 August)*. <https://cdn.ifla.org/past-wlic/2010/149-doeerr-en.pdf>.
- Dritsou, Vicky, Agiatis Benardou, Gustavo Candela, Sally Chambers, Alba Irollo, and Toma Tasovac. 2024a. "Introduction to Cultural Heritage Data." DARIAH-Campus Training Module. Version 1.0.0. <https://hdl.handle.net/21.11159/019595c5-1c30-7296-a43b-730c4eb50180>.
- Dritsou, Vicky, Agiatis Benardou, Gustavo Candela, Sally Chambers, Alba Irollo, and Toma Tasovac. 2024b. "Introduction to Cultural Heritage Data Modelling—With a Focus on Europeana Data Model." DARIAH-Campus Training Module. Version 1.0.0. <https://hdl.handle.net/21.11159/019595c5-1c63-725d-9f06-3f57952db0e0>.
- Dritsou, Vicky, Agiatis Benardou, Gustavo Candela, Sally Chambers, Alba Irollo, and Toma Tasovac. 2024c. "Introduction to Europeana APIs." DARIAH-Campus Training Module. Version 1.0.0. <https://hdl.handle.net/21.11159/019595c5-1d3a-70b7-85ac-6b4b87e7f65f>.
- Ducatteeuw, Vincent, Daniel Biltereyst, Philippe Meers, Christophe Verbruggen, Dries Moreels, Julia Noordegraaf, Sally Chambers, et al. 2023. "Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium." *Journal of Open Humanities Data* 9 (1). <https://doi.org/10.5334/johd.91>.
- European Commission. 2020. "Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)." COM/2020/767 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>.
- Europeana Foundation. 2008. "Europeana.eu Dataset." Dataset. 20 November 2008. <https://data.europa.eu/data/datasets/http-data-europeana-eu-dataset-europeana>.
- Europeana Foundation. 2016. *Europeana Dataset Profile*. Europeana. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_profiles/EDM_Dataset_Profile_042016.pdf.
- Franklin, Michael, Alon Halevy, and David Maier. 2005. "From Databases to Dataspaces: A New Abstraction for Information Management." *ACM SIGMOD Record* 34 (4): 27–33. <https://doi.org/10.1145/1107499.1107502>.
- Freire, Nuno. 2020. "Specifying a Linked Data Dataset for Europeana and Aggregators." Zenodo. Version 0.2, May. <https://doi.org/10.5281/zenodo.3817314>.

- Freire, Nuno, Hugo Manguinhas, Antoine Isaac, and Valentine Charles. 2023. "Persistent Identifier Usage by Cultural Heritage Institutions: A Study on the European.eu Dataset." In *Linking Theory and Practice of Digital Libraries: 27th International Conference on Theory and Practice of Digital Libraries, TPDL 2023 (Zadar, Croatia, 26–29 September), Proceedings*, 341–348. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-031-43849-3_31.
- Gangemi, Aldo, Silvio Peroni, David Shotton, and Fabio Vitali. 2014. "The Publishing Workflow Ontology (PWO)." *Semantic Web* 8 (5): 703–718. <https://doi.org/10.3233/SW-160230>.
- Geburu, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna M. Wallach, Hal Daumé III, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92. <https://doi.org/10.1145/3458723>.
- Hynning, Victoria Van, Lauren Algee, Mason Jones, Carlynn Osborn, Trevor Owens, Lauren Seroka, and Abby Shelton. 2022. "By the People Crowdsourcing Datasets from the Library of Congress." *Journal of Open Humanities Data* 8 (0). <https://doi.org/10.5334/johd.67>.
- International Data Spaces Association. 2019. *Reference Architecture Model*. Version 3.0. April 2019. <https://internationaldataspaces.org/wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf>.
- International GLAM Labs Community. 2022. "Computational Access to Digital Collections." <https://glamlabs.io/computational-access-to-digital-collections/>.
- Lehmann, Jörg, and Sophie Schneider. 2024. "Metadata of the 'Alter Realkatalog' (ARK) of Berlin State Library (SBB)." Dataset. Zenodo. Version v2, August. <https://doi.org/10.5281/zenodo.13284442>.
- Mahey, Mahendra, Aisha Al-Abdulla, Sarah Ames, Paula Bray, Gustavo Candela, Caleb Derven, Milena Dobrev, McPherson, et al. 2019. *Open a GLAM Lab*. Doha, Qatar: International GLAM Labs Community (Book Sprint). <http://dx.doi.org/10.21428/16ac48ec.f54af6ae>.
- Niccolucci, Franco, Achille Felicetti, and Sorin Hermon. 2022. "Populating the Data Space for Cultural Heritage with Heritage Digital Twins." *Data* 7 (8): 105. <https://doi.org/10.3390/data7080105>.
- OECD. 2022. "Measuring the Value of Data and Data Flows." *OECD Digital Economy Papers*, no. 345, <https://doi.org/10.1787/923230a6-en>.
- Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke, and Stewart Varner. 2019a. "Final Report—Always Already Computational: Collections as Data." Zenodo. Version 1, 22 May. <https://doi.org/10.5281/zenodo.3152935>.
- Padilla, Thomas, Laurie Allen, Hannah Frost, Sarah Potvin, Elizabeth Russey Roke, and Stewart Varner. 2019b. "Santa Barbara Statement on Collections as Data—Always Already Computational: Collections as Data." Zenodo. Version 2, 20 May. <https://doi.org/10.5281/zenodo.3066209>.
- Padilla, Thomas, Hannah Scates Kettler, Stewart Varner, and Yasmeen Shorish. 2023. "Vancouver Statement on Collections as Data." Zenodo. Version v3, 13 September. <https://doi.org/10.5281/zenodo.8342171>.
- Puren, Marie, Charles Riondet, Klaus Illmayer, Laurent Romary, Nicolas Larrousse, Laure Barbot, Matej Ďurčo, and Edward Gray. 2023. "How to Create a Workflow in the SSH Open Marketplace?" SSH Open Marketplace. <https://marketplace.sshopencloud.eu/workflow/hmGpmv>.
- Royal Library of Belgium. 2020. "DATA-KBR-BE. Facilitating Data-level Access to KBR's Digitised and Born-digital Collections for Digital Humanities Research." KBR. <https://www.kbr.be/en/projects/data-kbr-be/>.
- Russell, Nick, Arthur Ter, Wil Aalst, and Nataliya Mulyar. 2006. *Workflow Control-Flow Patterns: A Revised View*. BPM Reports, vol. 0622. BPM Center. <https://pure.tue.nl/ws/portalfiles/portal/2456784/710796038377571.pdf>.
- Sherratt, Tim. 2021. "GLAM Workbench." Zenodo. Version v1.0.0, 27 October. <https://doi.org/10.5281/zenodo.5603060>.

- Shiri, Ali, Liz Woolcott, Santi Thompson, Ayla Stein Kenfield, Elizabeth Joan Kelly, Joyce Chapman, and Derrick Jefferson. 2023. "A Conceptual Framework for the Design of Digital Repository Interfaces Supporting Digital Content Reuse Assessment." In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2023 (Santa Fe, USA, June 26–30)*, 30–34. IEEE. <https://doi.org/10.1109/JCDL57899.2023.00014>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (March). <https://doi.org/10.1038/sdata.2016.18>.
- World Wide Web Consortium. 2024. *Data Catalog Vocabulary (DCAT)*. Edited by Riccardo Albertoni, David Browning, Simon JD Cox, Alejandra Gonzalez Beltran, Andrea Perego, and Peter Winstanley. W3C. Version 3. Recommendation 22 August 2024. <https://www.w3.org/TR/vocab-dcat-3/>.
- Wu, Mei-Shin, Nathanael E. Schweikhard, Timotheus A. Bodt, Nathan W. Hill, and Johann-Mattis List. 2020. "Computer-Assisted Language Comparison: State of the Art." *Journal of Open Humanities Data* 6 (1). <https://doi.org/10.5334/johd.12>.

ANNEX

DCAT-AP specifies mandatory, recommended and optional data elements. It is mandatory to provide an instance of a `dcat:Catalogue`, which must be accompanied by an instance of a `foaf:Agent` that represents the publisher of the catalogue. The dataset(s) comprising the catalogue must be represented as an instance (or instances) of `dcat:Dataset`.

This appendix presents the DCAT-AP classes used for mapping the metadata about Europeana's datasets. The mapping is made from the datasets metadata that the Europeana Foundation maintains according to the EDM Dataset Profile (Europeana Foundation 2016). Note that instances of some classes carry very little information (for example `dct:LicenseDocument`): they are added for basic compliance with the validation rules at data.europa.eu, which is not yet able to handle "implicit" instances of these classes that would be used in the descriptions of the "main" classes (say, when a certain licence is used as the object of `dct:license` for an instance of `dcat:Catalogue`).

DCAT-AP and EDM reuse terms from various existing specifications. To facilitate readability, the classes and properties in this section are prefixed according to their namespaces, as follows:

- `adms`: <http://www.w3.org/ns/adms>
- `dcat`: <http://www.w3.org/ns/dcat#>
- `dcatap`: <http://data.europa.eu/r5r/>
- `dct`: <http://purl.org/dc/terms/>
- `edm`: <http://www.europeana.eu/schemas/edm/>
- `foaf`: <http://xmlns.com/foaf/0.1/>
- `rdfs`: <http://www.w3.org/2000/01/rdf-schema#>
- `skos`: <http://www.w3.org/2004/02/skos/core#>
- `xsd`: <http://www.w3.org/2001/XMLSchema#>

CLASS *DCAT:CATALOGUE*

Table 1 describes the class `dcat:Catalogue`.

CLASS *DCAT:DATASET*

Table 2 describes the class `dcat:Dataset` for the full dataset aggregated by the Europeana Foundation, while table 3 describes the class `dcat:Dataset` for individual datasets from CH institutions.

[Data.europa.eu](http://data.europa.eu) uses a controlled vocabulary for dataset categories, which are stated by `dcat:theme` properties. Table 4 shows the categories used and their corresponding URIs.

In the future, we may want to add references to Europeana's usage guidelines³⁷ and attribution pages.³⁸

37. <https://www.europeana.eu/en/rights/usage-guidelines-for-metadata>

38. <https://www.europeana.eu/en/rights/europeana-data-sources>

Table 1: Class `dcatalog:Catalogue`.

Class <code>dcatalog:Catalogue</code> .			
Mapping note:	This class is used for the root resource of Europeana's catalogue of datasets.		
URI:	http://data.europeana.eu/catalog		
Property	Range	Card.	Mapping note
<code>dcatalog:dataset</code>	<code>dcatalog:Dataset</code>	1..n	The URI http://data.europeana.eu/dataset-europeana that refers to the aggregated dataset of Europeana.eu and the DOI that is assigned by Zenodo to the datasets produced in the context of Saint George on a Bike (for example, https://doi.org/10.5281/zenodo.6984525). We considered including each of the individual datasets in Europeana.eu but have not pursued it.
<code>dct:description</code>	<code>rdfs:Literal</code>	1..n	The literal: "The catalogue of all cultural heritage datasets aggregated by Europeana"@en.
<code>dct:publisher</code>	<code>foaf:Agent</code>	1..n	The value must be an instance of a <code>foaf:Agent</code> describing the Europeana Foundation. See the mapping for <code>foaf:Agent</code> .
<code>dct:title</code>	<code>rdfs:Literal</code>	1..1	The literal "Europeana.eu Datasets"@en.
<code>foaf:homepage</code>	<code>foaf:Document</code>	0..1	The URI https://www.europeana.eu/ .
<code>dcatalog:keyword</code>	<code>rdfs:Literal</code>	0..n	The value: "culture"@en. The property may be repeated for translations of the keyword. The values should have an <code>xml:lang</code> tag.
<code>dct:language</code>	<code>dct:LinguisticSystem</code>	0..n	The Europeana catalogue may contain data in multiple languages, depending on the languages in use in the descriptions for the datasets, therefore no value should be provided for <code>dct:language</code> .
<code>dct:license</code>	<code>dct:LicenseDocument</code>	0..1	The value should be the URI for CC0, http://creativecommons.org/publicdomain/zero/1.0/ , described separately as a resource of type <code>dct:LicenseDocument</code> . See the mapping for <code>dct:LicenseDocument</code> .
<code>dct:issued</code>	<code>rdfs:Literal</code> typed as <code>xsd:date</code> or <code>xsd:dateTime</code>	0..1	Must be an <code>xsd:date</code> or an <code>xsd:dateTime</code> . We chose 1 March 2021 as the date the second iteration of MS7 was due and the approximate date of submission to data.europa.eu.
<code>dct:spatial</code>	<code>dct:Location</code>	0..n	The URI of the entity for Europe defined in the EU Vocabularies: http://publications.europa.eu/resource/authority/continent/EUROPE . See also the mapping for <code>dct:Location</code> .
<code>dcatalog:theme</code>	<code>skos:Concept</code>	0..n	The URI for the category "Education, Culture and Sport" of data.europa.eu's controlled vocabulary for dataset categories: http://publications.europa.eu/resource/authority/data-theme/EDUC . The list of data.europa.eu categories can be consulted in Table 4.
<code>dcatalog:themeTaxonomy</code>	<code>skos:ConceptScheme</code>	0..n	The URI of the Dataset Theme Vocabulary: http://publications.europa.eu/resource/authority/data-theme .
<code>dct:modified</code>	<code>rdfs:Literal</code> typed as <code>xsd:date</code> or <code>xsd:dateTime</code>	0..1	Must be an <code>xsd:date</code> or an <code>xsd:dateTime</code> . This should be the last date of modification of the DCAT-AP file.
<code>dct:creator</code>	<code>foaf:Agent</code>	0..n	The value must be the instance of a <code>foaf:Agent</code> describing the Europeana Foundation. See the mapping for <code>foaf:Agent</code> .

CLASS *DCAT:DISTRIBUTION*

Instances of `dcatalog:Distribution` should be provided by Europeana in two cases:

- For the whole dataset, the available distributions are Europeana's SPARQL endpoint, OAI-PMH API, data dumps (in RDF/XML and RDF Turtle syntax), Search API and Record API.
- For the individual datasets from CH institutions, Europeana makes data dumps available in RDF/XML and RDF Turtle.

Tables 5 and 6 present the mappings for the two cases separately.

CLASS *DCAT:DATASERVICE*

Several instances of `dcatalog:DataService` may be provided by Europeana to describe the APIs that provide public access to its data distributions. This section presents the mapping (see table 7) for the `dcatalog:DataService` instance that represents the SPARQL endpoint of Europeana.eu.

CLASS *FOAF:AGENT*

Instances of `foaf:Agent` must be provided by Europeana in two cases. First, to represent the Europeana Foundation, and second, to represent the CH institutions, in which case Europeana provides a description for individual (provider-level) datasets. Tables 8 and 9 present the mappings for the two cases separately.

Once we can attach organisation data to DCAT-AP records, we will revisit this mapping.

CLASS *DCT:LICENSEDOCUMENT*

Table 10 shows the class `dct:LicenseDocument`.

CLASS *SKOS:CONCEPT*

Table 11 shows the class `skos:Concept`.

Table 2: Class `dcat:Dataset`.

Class <code>dcat:Dataset</code> (for the dataset aggregated by the Europeana Foundation).			
Mapping note:	An instance of this class is used to represent the Europeana.eu aggregated dataset of cultural heritage digital objects. The class is mapped from <code>edm:Collection</code> defined in the Europeana Datasets Profile.		
URI:	http://data.europeana.eu/catalog		
Property	Range	Card.	Mapping note
<code>dct:description</code>	<code>rdfs:Literal</code>	1..n	The value: "A dataset containing metadata descriptions of cultural heritage digital objects from, and about, Europe. The dataset aggregates metadata from more than 4,000 cultural heritage institutions."@en The property may be repeated for providing translations. The values should have an <code>xml:lang</code> tag.
<code>dct:title</code>	<code>rdfs:Literal</code>	1..n	The value: "Europeana.eu aggregated dataset of cultural heritage digital objects"@en. The property may be repeated for translations of the title. The values should have an <code>xml:lang</code> tag.
<code>dcat:keyword</code>	<code>rdfs:Literal</code>	0..n	The value: "culture"@en. The property may be repeated for translations of the keyword. The values should have an <code>xml:lang</code> tag.
<code>dcat:distribution</code>	<code>dcat:Distribution</code>	0..n	One value should be created specifying the distribution via the SPARQL API of Europeana. See the mapping for <code>dcat:Distribution</code> .
<code>dct:publisher</code>	<code>foaf:Agent</code>	0..1	The value should be the instance of <code>foaf:Agent</code> representing the Europeana Foundation. See the mapping for <code>foaf:Agent</code> .
<code>foaf:page</code>	<code>foaf:Document</code>	0..n	Links to web pages and documents about the dataset. It should include links to the EDM documentation, and to the instructions for using all the Europeana APIs and data dumps described in the distributions. ^o
<code>dcat:contactPoint</code>	<code>vcard:Kind</code>	0..n	The value should be the URI for the Europeana Foundation. The <code>foaf:Agent</code> representing Europeana should also be typed as <code>vcard:Kind</code> .
<code>dct:spatial</code>	<code>dct:Location</code>	0..n	The URI of the entity for Europe defined in the EU Vocabularies: http://publications.europa.eu/resource/authority/continent/EUROPE . See also the section for <code>dct:Location</code> .
<code>dct:temporal</code>	<code>dct:PeriodOfTime</code>	0..n	The value should be the instance of <code>dct:PeriodOfTime</code> representing a time span with the earliest date associated, with an object in Europeana.eu as the start date, and the current year as the end date. See the mapping for <code>dct:PeriodOfTime</code> .
<code>dcat:theme</code>	<code>skos:Concept</code>	0..n	The URI for the category "Education, Culture and Sport" of data.europa.eu's controlled vocabulary for dataset categories: http://publications.europa.eu/resource/authority/data-theme/EDUC . The list of data.europa.eu categories can be consulted in Table 4.
<code>dct:creator</code>	<code>foaf:Agent</code>	0..n	The value should be the instance of <code>foaf:Agent</code> representing the Europeana Foundation. See the mapping for <code>foaf:Agent</code> .
<code>dct:conformsTo</code>	<code>dct:Standard</code>	0..n	A value for the EDM schema, stated with the URI http://www.europeana.eu/schemas/edm/ .
<code>dct:issued</code>	<code>rdfs:Literal</code> typed as <code>xsd:date</code> or <code>xsd:dateTime</code>	0..1	Must be an <code>xsd:date</code> or an <code>xsd:dateTime</code> . We use the date of Europeana's official launch on 20 November 2008. ("2008-11-20"8sd:date)
<code>dct:modified</code>	<code>rdfs:Literal</code> typed as <code>xsd:date</code> or <code>xsd:dateTime</code>	0..1	Must be an <code>xsd:gYear</code> , an <code>xsd:date</code> or an <code>xsd:dateTime</code> (in the latest ingestion in data.europa.eu we used "2022"8sd:gYear). Future ingestions should use the current year.
<code>dct:accessRights</code>	<code>dct:RightsStatement</code>	0..1	The value should be the URI for public access, http://publications.europa.eu/resource/authority/access-right/PUBLIC , described separately as a resource of type <code>dct:RightsStatement</code> .

^o. At first the `foaf:page` statements were attached to the Distributions but were not rendered well on data.europa.eu. So in agreement with the portal support team we have moved these statements onto the Dataset instance. This may be revisited again later.

CLASS `DCT:LOCATION`

Table 12 shows the class `dct:Location`.

CLASS `DCT:MEDIATYPEOREXTENT`

Table 13 shows the class `dct:MediaTypeOrExtent`.

CLASS `SKOS:CONCEPTSCHEME`

Table 14 shows the class `skos:ConceptScheme`.

CLASS `DCT:PERIODOFTIME`

Table 15 shows the class `dct:PeriodOfTime`.

CLASS `FOAF:DOCUMENT`

Table 16 shows the class `foaf:Document`.

CLASS `DCT:STANDARD`

The `dct:Standard` class is the range defined by DCAT-AP for `dct:conformsTo` properties of datasets and distributions. DCAT-AP does not specify any properties to be used for instances of `dct:Standard`, therefore we only focus here on identifying the URIs to refer to the standards in use by Europeana. One of the uses of `dct:conformsTo` is to express the usage of the Europeana Data Model, or EDM, as the schema of the datasets. In this case, the URI of the EDM namespace must be used: <http://www.europeana.eu/schemas/edm/>. Another use

Table 3: Class `dcat:Dataset` (for individual datasets from CH institutions).

Class <code>dcat:Dataset</code> (for individual datasets from CH institutions).			
Mapping note:	This class is used to represent the datasets in Europeana.eu. The class is mapped from <code>edm:Collection</code> defined in the Europeana Datasets Profile.		
URI:	Should we decide to publish representations of individual datasets, the URIs should be assigned according to a pattern based on the Europeana "dataset name."		
Property	Range	Card.	Mapping note
<code>dct:description</code>	<code>rdfs:Literal</code>	1..n	Mapped from the <code>dct:description</code> . The property may be repeated for translations of the description. The values should have an <code>xml:lang</code> tag.
<code>dct:title</code>	<code>rdfs:Literal</code>	1..n	Mapped from <code>edm:datasetName</code> . The <code>edm:datasetName</code> does not provide a real, human-readable title, however. They are strings like "2022362_Ag_UK_CultureGrid_Royal_Museums_Greenwich", which cannot be considered to belong to any language (and thus should have the value for "zxx" – no linguistic content – in the <code>xml:lang</code> tag).
<code>dcat:keyword</code>	<code>rdfs:Literal</code>	0..n	The value: "culture"@en. The property may be repeated for translations of the keyword. The values should have an <code>xml:lang</code> tag.
<code>dcat:distribution</code>	<code>dcat:Distribution</code>	0..n	One value should be created specifying the file dump created by the Europeana Foundation for distributing the dataset. See the mapping for the <code>dcat:Distribution</code> class.
<code>dct:publisher</code>	<code>foaf:Agent</code>	0..1	The value should be the instance of <code>foaf:Agent</code> representing the Europeana Foundation.
<code>dcat:contactPoint</code>	<code>vcard:Kind</code>	0..n	The value should be the URI for the Europeana Foundation. The <code>foaf:Agent</code> representing the Europeana Foundation should also be typed as <code>vcard:Kind</code> .
<code>dct:creator</code>	<code>foaf:Agent</code>	0..n	Mapped from the <code>edm:dataProvider</code> and represented as a <code>foaf:Agent</code> . See the mapping for the <code>foaf:Agent</code> class. In order to recognise contributions from Europeana's aggregators we would have to create <code>dct:creator</code> statements for the <code>edm:provider</code> and <code>edm:intermediateProvider</code> organisations. However, <code>dct:creator</code> is not a great fit for their role. If DCAT-AP would use other roles, like <code>dct:contributor</code> , then we could switch to using it. We could also use <code>dct:publisher</code> as a better (but not perfect) fit, if DCAT-AP would allow us to list more than one per dataset.
<code>dct:spatial</code>	<code>dct:Location</code>	0..n	The EDM Datasets Profile features an <code>edm:country</code> property that could be an interesting candidate for <code>dct:spatial</code> . But in effect, it is used to indicate the country of a dataset's provider and is therefore a rather poor fit for the spatial coverage of a dataset.
<code>dct:conformsTo</code>	<code>dct:Standard</code>	0..n	A value for the EDM schema, stated with the URI http://www.europeana.eu/schemas/edm/ .
<code>dcat:theme</code>	<code>skos:Concept</code>	0..n	One value should be provided using data.europa.eu's controlled vocabulary for dataset categories. Consult the table at the end of this section. Since the Datasets Profile does not contain any property for categorising the dataset, all datasets should have the most generally applicable category in Europeana value for <code>dcat:theme</code> : http://publications.europa.eu/resource/authority/data-theme/EDUC . There might still be ways to automatically assign categories, but this requires more investigation.
<code>dct:accessRights</code>	<code>dct:RightsStatement</code>	0..1	The value should be the URI for public access, http://publications.europa.eu/resource/authority/access-right/PUBLIC , described separately as a resource of type <code>dct:RightsStatement</code> .

Table 4: Controlled vocabulary for the dataset categories used on data.europa.eu.

Category	URI
Agriculture, fisheries, forestry and food	http://publications.europa.eu/resource/authority/data-theme/AGRI
Education, culture and sport	http://publications.europa.eu/resource/authority/data-theme/EDUC
Energy	http://publications.europa.eu/resource/authority/data-theme/ENER
Environment	http://publications.europa.eu/resource/authority/data-theme/ENVI
Transport	http://publications.europa.eu/resource/authority/data-theme/TRAN
Science and technology	http://publications.europa.eu/resource/authority/data-theme/TECH
Economy and finance	http://publications.europa.eu/resource/authority/data-theme/ECON
Population and society	http://publications.europa.eu/resource/authority/data-theme/SOCI
Health	http://publications.europa.eu/resource/authority/data-theme/HEAL
Government and public sector	http://publications.europa.eu/resource/authority/data-theme/GOVE
Regions and cities	http://publications.europa.eu/resource/authority/data-theme/REGI
Justice, legal system and public safety	http://publications.europa.eu/resource/authority/data-theme/JUST
International issues	http://publications.europa.eu/resource/authority/data-theme/INTR
Provisional data	http://publications.europa.eu/resource/authority/data-theme/OP_DATPRO

of `dct:conformsTo` is to express the conformity of the distributions of the Europeana.eu datasets via the SPARQL endpoint, the OAI-PMH API or data dumps. In this case, for example, the URI of the SPARQL query specification must be used: <http://www.w3.org/TR/sparql11-query/>.

Table 5: Class `dcate:Distribution` (for the distributions for the whole Europeana.eu dataset via SPARQL endpoint, OAI-PMH API, data dumps in RDF/XML and RDF Turtle syntax, the Search API and the Record API).

Class <code>dcate:Distribution</code> (for the distributions for the whole Europeana.eu dataset via SPARQL endpoint, OAI-PMH API, data dumps in RDF/XML and RDF Turtle syntax, the Search API and the Record API).			
Mapping note:	These instances of <code>dcate:Distribution</code> are used to represent the access to the complete aggregated dataset of Europeana.eu via Europeana's SPARQL endpoint, OAI-PMH API and data dumps. Note that the values for some of the properties will differ, depending on the distribution being represented.		
URI:	For the time being these instances are going to be treated as unnamed individuals (blank nodes).		
Property	Range	Card.	Mapping note
<code>dcate:accessURL</code>	<code>rdfe:Resource</code>	1..n	The distribution-specific access URL such as https://sparql.europeana.eu/ . Note that it should be an HTTPS URL that leads to a user-readable web page, since it is used by data.europeana.eu to lead users to the distribution.
<code>dcate:availability</code>	<code>skos:Concept</code>	0..1	The value must be the URI defined by DCAT-AP: http://data.europeana.eu/r5r/stable .
<code>dct:title</code>	<code>rdfe:Literal</code>	0..n	The value "Europeana.eu dataset (<DISTRIBUTION NAME>)"@en.
<code>dct:description</code>	<code>rdfe:Literal</code>	0..n	The value should be a short description of the distribution. Must be language-tagged and may be repeated for descriptions in several languages.
<code>dct:format</code>	<code>dct:MediaTypeOrExtent</code>	0..1	Should always contain one value consisting of a URI from the EU Vocabularies File Type Named Authority List. ^a
<code>dct:license</code>	<code>dct:LicenseDocument</code>	0..1	The value should be the URI for CC0, http://creativecommons.org/publicdomain/zero/1.0/ , described separately as a resource of type <code>dct:LicenseDocument</code> . See the mapping for <code>dct:LicenseDocument</code> .
<code>dct:rights</code>	<code>dct:RightsStatement</code>	0..1	The value should be the URI for CC0, http://creativecommons.org/publicdomain/zero/1.0/ , described separately as a resource of type <code>dct:RightsStatement</code> .
<code>dcate:accessService</code>	<code>dcate:DataService</code>	0..n	The URI of the <code>dcate:DataService</code> representing the SPARQL endpoint, OAI-PMH API, data dumps, Search API and Record API of Europeana.eu. See the mapping for <code>dcate:DataService</code> .
<code>dct:conformsTo</code>	<code>dct:Standard</code>	0..n	Two values should be provided for <code>dct:conformsTo</code> : a value for the EDM schema, stated with the URI http://www.europeana.eu/schemas/edm/ ; and another value for distribution-specific standards. For example, SPARQL with the URI http://www.w3.org/TR/sparql11-query/ . ^b
<code>dcate:mediaType</code>	<code>dct:MediaTypeOrExtent</code>	0..1	The IANA URIs: <ul style="list-style-type: none"> for the data dumps: https://www.iana.org/assignments/media-types/application/rdf+xml or https://www.iana.org/assignments/media-types/text/turtle for the SPARQL endpoint: https://www.iana.org/assignments/media-types/text/strings for the OAI-PMH access: https://www.iana.org/assignments/media-types/application/xml for the Search and Record APIs: https://www.iana.org/assignments/media-types/application/json
<code>dcate:packageFormat</code>	<code>dct:MediaType</code>	0..1	For the data dump distributions, the URI: https://www.iana.org/assignments/media-types/application/zip .
<code>adms:status</code>	<code>skos:Concept</code>	0..1	The value must be a URI defined by ADMS, such as http://purl.org/adms/status/UnderDevelopment or http://purl.org/adms/status/Completed .

a. <https://publications.europeana.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/file-type>

b. We could have both the query language and the protocol URIs, see the ongoing discussion at <https://github.com/w3c/dxwg/issues/1211>, which at the time of writing has not been concluded.

Table 6: Class `dcate:Distribution` (two for each individual Europeana.eu dataset: one in RDF/XML and one in RDF Turtle syntax).

Class <code>dcate:Distribution</code> (two for each individual Europeana.eu dataset: one in RDF/XML and one in RDF Turtle syntax).			
Mapping note:	This class is used to represent the file dumps and other relevant distributions that Europeana publishes for each of its datasets.		
URI:	N/A (the option of publishing individual datasets has not been pursued so far).		
Property	Range	Card.	Mapping note
<code>dcate:accessURL</code>	<code>rdfe:Resource</code>	1..n	The URL: https://europeana.atlassian.net/wiki/spaces/EF/pages/2324463617/Dataset+download+and+OAI-PMH+service . Note that it should be an HTTPS URL that leads to a user-readable web page, since it is used by data.europeana.eu to lead users to the distribution.
<code>dcate:availability</code>	<code>skos:Concept</code>	0..1	The value must be the URI defined by DCAT-AP: http://data.europeana.eu/r5r/available .
<code>dct:description</code>	<code>rdfe:Literal</code>	0..n	The value: "A ZIP archive containing the dataset in EDM"@en.
<code>dct:format</code>	<code>dct:MediaTypeOrExtent</code>	0..1	The URIs: http://publications.europeana.eu/resource/authority/file-type/RDF_TURTLE or http://publications.europeana.eu/resource/authority/file-type/RDF_XML .
<code>dct:license</code>	<code>dct:LicenseDocument</code>	0..1	The value should be the URI for CC0, http://creativecommons.org/publicdomain/zero/1.0/ , described separately as a resource of type <code>dct:LicenseDocument</code> . See the mapping for <code>dct:LicenseDocument</code> .
<code>dct:rights</code>	<code>dct:RightsStatement</code>	0..1	The value should be the URI for CC0, http://creativecommons.org/publicdomain/zero/1.0/ , described separately as a resource of type <code>dct:RightsStatement</code> .
<code>dcate:byteSize</code>	<code>rdfe:Literal</code> typed as <code>xsd:decimal</code>	0..1	Automatically generated.
<code>spdx:checksum</code>	<code>spdx:Checksum</code>	0..1	Automatically generated.
<code>dcate:downloadURL</code>	<code>rdfe:Resource</code>	0..n	URL pattern: <a href="ftp://download.europeana.eu/dataset/XML/<dataset_id>.zip">ftp://download.europeana.eu/dataset/XML/<dataset_id>.zip .
<code>dct:conformsTo</code>	<code>dct:Standard</code>	0..n	A value for the EDM schema, stated with the URI http://www.europeana.eu/schemas/edm/ .
<code>dcate:mediaType</code>	<code>dct:MediaType</code>	0..1	The IANA URIs: https://www.iana.org/assignments/media-types/application/rdf+xml or https://www.iana.org/assignments/media-types/text/turtle .
<code>dcate:packageFormat</code>	<code>dct:MediaType</code>	0..1	The URI: https://www.iana.org/assignments/media-types/application/zip .
<code>adms:status</code>	<code>skos:Concept</code>	0..1	For all datasets published by Europeana the value must be the URI defined by ADMS: http://purl.org/adms/status/Completed .
<code>dct:modified</code>	<code>rdfe:Literal</code> typed as <code>xsd:date</code> or <code>xsd:dateTime</code>	0..1	Automatically generated.

Table 7: Class `dcate:DataService` (for the Europeana.eu SPARQL API, the OAI-PMH API, the Search API and the Record API).

Class <code>dcate:DataService</code> (for the Europeana.eu SPARQL API, the OAI-PMH API, the Search API and the Record API).			
Mapping note:	This class is used to represent the Europeana.eu SPARQL API, the OAI-PMH API, the Search API and the Record API.		
URI:	For the time being, instances of <code>dcate:DataService</code> are going to be treated as unnamed individuals (blank nodes) because each service corresponds to one dataset.		
Property	Range	Card.	Mapping note
<code>dcate:endpointURL</code>	<code>rdfs:Resource</code>	0..n	The service-specific URL, such as https://sparql.europeana.eu or https://api.europeana.eu/record/ .
<code>dct:title</code>	<code>rdfs:Literal</code>	1..n	One of the literals: "Europeana SPARQL API"@en, "Europeana OAI-PMH API", "Europeana Search API", "Europeana Record API".
<code>dcate:servesDataset</code>	<code>dcate:Dataset</code>	0..n	The URI of the Europeana.eu dataset: http://data.europeana.eu#dataset-europeana .
<code>dct:description</code>	<code>rdfs:Literal</code>	1..n	The value should be a short description of the data service. Must be language-tagged and may be repeated for descriptions in several languages.
<code>dcate:endpointDescription</code>	<code>rdfs:Resource</code>	0..n	Should contain the URL of a web page describing the services available via the end-points, including their operations, parameters, etc. Typically, these will point to a web page in the Europeana Knowledge base, such as https://europeana.atlassian.net/wiki/spaces/EF/pages/2385870903/SPARQL+API+Documentation or https://europeana.atlassian.net/wiki/spaces/EF/pages/2385674279/Record+API+Documentation .

Table 8: Class `foaf:Agent` (for the Europeana Foundation).

Class <code>foaf:Agent</code> (for the Europeana Foundation).			
Mapping note:	This class is used to represent the Europeana Foundation.		
URI:	http://data.europeana.eu/organization/1482250000001617026		
Property	Range	Card.	Mapping note
<code>foaf:name</code>	<code>rdfs:Literal</code>	1..n	The property should have two literal values: "Europeana Foundation"@en and "Stichting Europeana"@nl. It may be repeated for translations of the name. The values should have an <code>xml:lang</code> tag.
<code>dct:type</code>	<code>skos:Concept</code>	0..1	The value should be the URI for non-profit organisation defined by ADMS: http://purl.org/adms/publictype/NonProfitOrganisation .

Table 9: Class `foaf:Agent` (for the CH institutions).

Class <code>foaf:Agent</code> (for the CH institutions).			
Mapping note:	This class is used to represent the CH institutions. These instances of <code>foaf:Agent</code> are used in properties of individual Europeana.eu datasets. Instances of <code>foaf:Agent</code> should also be created for providers and intermediate providers in this case.		
URI:	Europeana is in the process of minting and publishing identifiers for its data providers (via the Entity API), which we will use here when individual datasets are described.		
Property	Range	Card.	Mapping note
<code>foaf:name</code>	<code>rdfs:Literal</code>	1..n	The value corresponds to either the property <code>edm:dataProvider</code> , <code>edm:provider</code> or <code>edm:intermediateProvider</code> that give rise to the creation of the Agent resource. The property may be repeated for translations of the name. The values should have an <code>xml:lang</code> tag.

Table 10: Class `dct:LicenseDocument`.

Class <code>dct:LicenseDocument</code> .			
Mapping note:	This class is used to represent the licences of the Europeana catalogue and of the datasets' distributions it lists. It is CC0, therefore the type of the resource shall be "public domain" in the classification suggested by DCAT-AP (see below).		
URI:	Since all data distributed by Europeana is made available under CC0, the URI must be http://creativecommons.org/publicdomain/zero/1.0/ .		
Property	Range	Card.	Mapping note
<code>dct:type</code>	<code>skos:Concept</code>	0..n	The URI of the public domain licence defined by ADMS: http://purl.org/adms/licencetype/PublicDomain .

Table 11: Class `skos:Concept`.

Class <code>skos:Concept</code> .			
Mapping note:	This class is the range defined by DCAT-AP for properties like <code>dct:type</code> and <code>dcate:theme</code> . Even when the values of these properties are URIs from the vocabularies supported in DCAT-AP, the metadata must provide some minimal data about the <code>skos:Concept</code> .		
URI:	The URI used in a value of <code>dct:type</code> or <code>dcate:theme</code> property.		
Property	Range	Card.	Mapping note
<code>skos:inScheme</code>	<code>skos:ConceptScheme</code>	1..n	The URI of the vocabulary where the <code>skos:Concept</code> is included.
<code>skos:prefLabel</code>	<code>Literal</code>	1..n	The preferred lexical label for a resource, in a given language.

Table 12: Class `dct:Location`.

Class <code>skos:Concept</code> .			
Mapping note:	The <code>dct:Location</code> class is the range defined by DCAT-AP for <code>dct:spatial</code> properties of catalogues and datasets. Even when the values of these properties are URIs from the EU Vocabularies Named Authority Lists used for continents, countries and places, the metadata must provide the URI of the vocabulary.		
URI:	The URI used in a value of <code>dct:spatial</code> property.		
Property	Range	Card.	Mapping note
<code>skos:inScheme</code>	<code>skos:ConceptScheme</code>	1..n	The URI of the vocabulary where the <code>dct:Location</code> is included, which is typically one of the EU Vocabularies' Named Authority Lists used for continents, countries and places.

Table 13: Class `dct:MediaTypeOrExtent`.

Class <code>dct:MediaTypeOrExtent</code> .			
Mapping note:	The <code>dct:MediaTypeOrExtent</code> class is the range defined by DCAT-AP for <code>dct:format</code> properties of distributions. Even when the values of these properties are URIs from the vocabularies supported in DCAT-AP, the metadata must provide the URI of the vocabulary.		
URI:	The URI used in a value of a <code>dct:format</code> property.		
Property	Range	Card.	Mapping note
<code>skos:inScheme</code>	<code>skos:ConceptScheme</code>	1..n	The URI of the vocabulary where the <code>dct:MediaTypeOrExtent</code> is included. DCAT-AP specifies that the EU Vocabularies File Type Named Authority List must be used for <code>dct:format</code> , therefore the value should be the URI http://publications.europa.eu/resource/authority/file-type .

Table 14: Class `skos:ConceptScheme`.

Class <code>skos:ConceptScheme</code> .			
Mapping note:	The <code>skos:ConceptScheme</code> class is used only for representing one of the vocabularies supported by DCAT-AP - Data Theme. This vocabulary must be defined in the DCAT-AP file in order to pass the DCAT-AP validation of data.europa.eu . The Europeana DCAP-AP metadata uses this vocabulary for values of <code>dcat:themeTaxonomy</code> properties of catalogues and <code>dcat:theme</code> properties of datasets.		
URI:	The URI of the Data Theme vocabulary – http://publications.europa.eu/resource/authority/data-theme .		
Property	Range	Card.	Mapping note
<code>dct:title</code>	Literal	1..n	The literal "Data Theme" with an "en" language tag.

Table 15: Class `dct:PeriodOfTime`.

Class <code>dct:PeriodOfTime</code> .			
Mapping note:	The <code>dct:PeriodOfTime</code> class is used for representing the time spans for <code>dct:temporal</code> properties of the datasets.		
URI:	These instances are going to be treated as unnamed individuals (blank nodes).		
Property	Range	Card.	Mapping note
<code>dcat:startDate</code>	Literal	0..n	The literal "-542000000" typed as <code>xsd:gYear</code> .
<code>dcat:endDate</code>	Literal	0..n	A literal with the current year (i.e. "2022") typed as <code>xsd:gYear</code> .

- a. Within the numerous archaeological items in Europeana.eu, the earliest date we were able to find was "-542000000/-251000000". It is present in a `dct:temporal` property of the following object: https://www.europeana.eu/en/item/2064135/Museu_ProvidedCHO_museum_digital_507__technical_number_

Table 16: Class `foaf:Document`.

Class <code>foaf:Document</code> .			
Mapping note:	The <code>foaf:Document</code> class is used to describe web pages and documents about the Europeana.eu dataset, its data services and distributions.		
URI:	The URL of the web page or document.		
Property	Range	Card.	Mapping note
<code>dct:title</code>	Literal	1..n	A language-tagged literal containing the title of the web page or document.
<code>dct:description</code>	Literal	1..n	The value should be a short description of the web page or document. Must be language-tagged and may be repeated for descriptions in several languages.
<code>schema:url</code>	<code>rdf:Resource</code>	0..n	The URL of the web page or document.
<code>dct:format</code>	<code>dct:MediaTypeOrExtent</code>	0..1	Should always contain one value consisting of a URI from the EU Vocabularies File Type Named Authority List. ^a In the case of HTML web pages, the value is http://publications.europa.eu/resource/authority/file-type/HTML .
<code>dcat:mediaType</code>	<code>dct:MediaType</code>	0..1	Must be an IANA URI which, in the case of HTML web pages, is http://www.iana.org/assignments/media-types/text/html .
<code>dct:type</code>	<code>skos:Concept</code>	0..1	Should always contain one value consisting of a URI from the EU Vocabularies Documentation Type Authority List, ^b such as http://publications.europa.eu/resource/authority/documentation-type/DOCUMENTATION_MAIN .

a. <https://publications.europa.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/file-type>

b. <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/documentation-type>