f

fRamework for safE, opEn, collaboratiVe

And incLUsive digitisAtion and managemenT

DELIVERABLE 1.1:

Comprehensive guide of the benefits, opportunities, risks and gaps in the management of cultural heritage digitisation

Work Package: 1

LEAD BENEFICIARY:

European Fashion Heritage Association (EFHA)

Delivery Date: 28.06.2024

Document Sheet

Project acronym	**REEVALUATE**
Project full title	Framework for safe, open, collaborative and inclusive digitization and management of cultural heritage
Programme	Horizon Europe
Topic	HORIZON-CL2-2023-HERITAGE-01-03
Type of Action	HORIZON-Research and Innovation Actions
Grant Agreement	101132389
Start day	1 January 2024
Duration	36 months

Document Information

Deliverable number	D1.1
Deliverable name	Comprehensive guide of the benefits, opportunities, risks and gaps in the management of cultural heritage digitisation
Lead beneficiary	EFHA
WP	1
Related task(s)	T1.1
Type	Report
Reviewers (Organisation)	CERTH, FS
Delivery date	28.06.2024
Main author(s)	Maria Drabczyk, Anastasia Dimou, Jasper de Koning, Francesca Mafredini, Jiri Svorc, Ruben Peeters
Contributor(s)	Marco Rendina

Dissemination level

PU	Public	X

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Document history

Version	Date	Changes	Reviewer/Contributor
0.1	07.06.2024	initial draft	Maria Drabczyk, Francesca Manfredini (EFHA)
0.2	12.06.2024	partner contributions	Anastasia Dimou (KU Leuven), Maria Drabczyk (EFHA), Jasper de Koning (Arthur), Francesca Mafredini (EFHA), Marco Rendina (EFHA), Jiri Svorc (ARTHUR), Ruben Peeters (KU Leuven)
0.3	14.06.2024	text integration and editing	Maria Drabczyk, Francesca Manfredini (EFHA)
0.4	27.06.2024	internal review	Babis Kyfonidis (CERTH) Anna-Maria Feneri, David Dokic (From-Scratch Design Studio)
0.5	27.06.2024	Integration of reviewers' comments	Maria Drabczyk (EFHA)
1	28.06.2024	Final draft	

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Table of Contents

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Abbreviations

Abbreviations	Full name
ACA	Resources for research or academic use
AI	Artificial Intelligence
APIs	Application Programming Interfaces
ARK	Archival Resource Key
CC0	Creative Commons Zero Universal Public Domain Dedication
CC-BY-SA	Creative Commons (CC) Attribution (BY) Share-alike (SA)
CDMI	Component Metadata Infrastructure
CH	Cultural Heritage
CHI	Cultural Heritage Institution
CLARIN	Common Language Resources and Technology Infrastructure
CMDI	Component Metadata Infrastructure
CMO	Collective Management Organisations
D4COLLECT	DARIAH working group on Combining Language Learning with Crowdsourcing Techniques
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DEA	Data Exchange Agreement

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

DGA	Data Governance Act
DHCR	Digital Humanities Course Registry
DOI	Digital Object Identifier
EDM	Europeana Data Model
ERIC	European Research Infrastructure Consortium
GLAM	Galleries, Libraries, Archives, Museums
IDS	International Data Spaces
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
LOD	Linked Open Data
MIMO	Musical Instrument Museums Online
NLP	Natural Language Processing
ODD	Open Data Directive
OIDC	OpenID Connect
ORCID	Open Researcher and Contributor ID
OWD	Orphans Work Directive
PDP	Policy Decision Point
PID	Persistent identifier

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

PUB	Resources publicly or openly available
RES	Resources restricted to individual use
SAML	Security Assertion Markup Language
SSO	Single Sign On
TOS	Terms of service
URN	Uniform Resource Name
WP	Work Package
XAI	Explainable AI

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Executive summary

The digitisation of cultural heritage, in theory, allows Cultural Heritage Institutions (CHIs) to make their collections more accessible to the public. Digitised and digital collections have the potential to be shared and utilised for research, education, and entertainment. Online access can help strengthen public engagement with cultural heritage, making it more relevant to contemporary audiences and turning them into active participants and co-creators working with digital artefacts, often alongside CHIs. These collaborations highlight the value of digital heritage to European society.

This document, analyses through a literature review the topic of digitisation management in cultural heritage in Europe, focusing on six macro areas: access and engagement, technology, legal and policy considerations, ethical considerations, skills and competencies, and sustainability. These areas are identified as essential for addressing and deciphering opportunities, risks and gaps in digitisation management that impact collaboration, creative reuse, and the promotion of democratic and inclusive prioritisation and contextualisation of digital cultural heritage.

An increase in participatory practices has been detected in the cultural sector, leading to more engagement with digital heritage. Digitisation and subsequent access to digitised collections allow various communities representing different stakeholder profiles (local communities, creatives, educators, researchers, technology entrepreneurs, and providers) to explore the potential and value of digital artefacts

Functioning in a digital environment—along with the new technologies, regulations and policies it introduces, as well as the inclusion of new voices in their operations—requires CHI professionals to develop a new mindset and acquire skills that enable them to embrace this emerging realm.

Nevertheless, digitisation, and especially digitisation leading to access, still presents numerous challenges and potential risks. A major challenge is the lack of comprehensive digitisation frameworks and policies, especially on national and regional levels, that provide a methodology and tools for CHIs to properly prioritise, contextualise and manage the lifecycle of their collections. Issues such as the loss of original context, copyright challenges and mismanagement of intellectual property rights (IPR), misuse of digitised Cultural Heritage (CH) objects, ownership risks, digital preservation risks, and organisational challenges (including sustainability, digital sobriety, and skills) remain significant hurdles in cultural heritage digitisation management. Additional challenges include long-term digital preservation, storage and technological obsolescence and funding. There are also ethical considerations increasingly taken into account when discussing access to and reuse of digitised and digital-born collections, incorporating previously neglected voices, embracing inclusivity, equity, and democratic values.

All of these factors can either enable or hinder the potential of digital heritage in society, affecting collaborations, innovation, creative reuse, and democratisation of cultural heritage organisations and their practices.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

1 Introduction

REEVALUATE is a Horizon Europe project designed to address the complex challenges of Cultural Heritage (CH) digitisation management with the goal of creating significant societal value and fostering deeper public engagement with cultural heritage. The project aims to create a holistic, modular framework that enhances collaboration, promotes creative reuse, and ensures democratic and inclusive prioritisation and contextualisation of CH artefacts. This framework will guide each stage of a digitised artefact's life cycle, namely prioritisation, contextualisation, storage, collaboration, and reuse, through innovative technological tools (enablers).

Central to REEVALUATE is the active participation of stakeholders. The project follows a Human-Centred Design approach, ensuring that the tools and framework developed are based on the real needs of stakeholders. AI techniques will be employed to assist manual contextualisation, improve prioritisation, visualise creative reuse ideas, propose collaborations, and validate the context of reused artefacts. The framework will also address secure storage and intellectual property rights management through CH object tokenization and smart contracts.

The REEVALUATE project will offer a standardised semantic representation of artefact metadata to improve accessibility and discoverability, while preventing misuse. Designed to be flexible, the enablers can be used independently or as an integrated solution, tailored to the specific needs of CH institutions.

REEVALUATE will validate its framework through three real-world pilot cases across different sectors—gaming, fashion, and advertising—and regions in Europe. The project will develop sustainable business models and policy recommendations to ensure long-term adoption and impact.

1.1 Work Package Relevance

WP1, "Review of the Current State of Digitisation of Cultural Heritage in Europe," focuses on a comprehensive evaluation of the digitisation status of cultural heritage across the EU. This involves identifying the potential benefits, opportunities, risks, and gaps associated with digitisation projects. A key aspect of this work package is assessing user requirements and needs, and developing specific usage cases for the REEVALUATE system. This approach ensures that the selected artefacts for digitisation, the formats used, and the manageability of the framework are practically applicable in the creative industries.

By the end of this process, user personas will be created, need lists for each user group will be compiled, and usage scenarios and types of digitised artefacts will be identified, along with their creative possibilities. The outcomes of this work package are essential for developing the methodological framework of REEVALUATE, ensuring it is efficient and easily applicable with the participation of all interested parties.

1.2 Purpose of this deliverable

The current document aims to assess the benefits, risks, and gaps in the digitisation of cultural heritage artefacts from a broader, high-level perspective, drawing on a selection of European literature, including policy documents, project reports, and scholarly articles. Its purpose is to help identify current approaches and practices in digitisation management to explore potential opportunities. By conducting a critical literature review, it seeks to offer a comprehensive guide that summarises and prioritises key elements of the digitisation management framework.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

The findings of this research phase will support further qualitative research in the project, implementation of the foreseen pilots, as well as the policy work of REEVALUATE.

With this deliverable, and in line with the project's objectives, REEVALUATE attempts to provide a comprehensive, actionable, and modular framework based on extensive multidirectional research for better management and use of CH artefacts' digitisation.

2 Overview of the subject under consideration

One of the main benefits of digitising cultural heritage is that, in theory, it allows Cultural Heritage Institutions to make their collections more accessible to the public. These digital collections can be shared and utilised for research, education, or entertainment. This online access helps strengthen public engagement with cultural heritage, making it more relevant to contemporary audiences. Specifically, by digitising artefacts, manuscripts, and other CH assets, institutions can create digital versions that can be viewed online and support digital interactions. Besides increasing access, digitization also improves the preservation of cultural heritage materials. Digital copies protect original artefacts from damage caused by handling, exposure to light, and pollution. They also help preserve materials and CH sites at risk of decay, destruction, or anthropogenic threats such as uncontrolled tourism development, urbanisation, war, or conflict.

For these reasons, the Cultural Heritage sector is under increasing pressure to digitise its collections. However, digitisation, and especially digitisation which leads to a wider access, still presents a number of challenges and potential risks and challenges. A major challenge is the lack of a comprehensive digitisation framework that will provide a methodology and tools for CHIs to properly prioritise, contextualise and manage the lifecycle of their collections. Issues such as the loss of original context, mismanagement of intellectual property rights (IPR), misuse of digitised CH objects, ownership risks linked to unclear legal status of works, preservation risks, stakeholder collaboration, and organisational challenges (including sustainability, digital sobriety, and new skills linked to digital management) remain significant hurdles in cultural heritage digitization management. Additional challenges include digital storage, technological obsolescence, and ethical considerations. These challenges and risks hinder the potential of digital heritage in society and need to be resolved in order to empower collaborations, innovation, creative reuse, and the promotion of democratic/inclusive prioritisation & contextualization.

REEVALUATE takes a broad view of the digitization by proposing a management framework, examining the lifecycle of both digitised and born-digital artefacts, including prioritisation, contextualization, storage, collaboration, and reuse. It aims to identify opportunities, risks, and gaps in current workflows and frameworks. The project focuses on involving all stakeholders, who are crucial at various stages of the digitization process, to achieve a greater impact in social, organisational, economic, and innovative contexts.

3 Methodological approach

3.1 Selection of literature review type

The exploration and analysis of the digitization eco-system and all the factors that influence it had to be studied in depth, in order for REEVALUATE to be able to safely and accurately design and implement its solution. The exploration and analysis were divided in two stages; 1) a literature review which covers a broader and more scientific/academic/regulatory view (this deliverable); and 2) a qualitative enquiry with interviews and surveys directly to the involved stakeholders (deliverable D1.2).

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

In order to select the most appropriate and effective, for the project's scope and resources, many types of literature reviews were identified and considered. However, the literature review type that has been chosen is the "critical review" as defined by Grant and Booth:

> "A critical review aims to demonstrate that the writer has extensively researched the literature and critically evaluated its quality. It goes beyond mere description of identified articles and includes a degree of analysis and conceptual innovation" and "an effective critical review presents, analyses and synthesises material from diverse sources". "There is no formal requirement to present methods of the search, synthesis and analysis explicitly." (Grant & Booth 2009)

This approach has been selected by the project team as it requires creativity, flexibility, and expert judgement, rather than systematicity, to explore current research and allows to propose a modified or new way forward, leveraging the team's expertise.

The strength of this methodology lies in its flexibility, which enables the advancement of understanding complex issues by appraising theory and evidence from various sources. Additionally, it allows researchers to continuously revise their interpretations of the problem, by using their perspectives to evaluate and interpret the uncovered literature. This approach is particularly useful in situations that require reviewers to apply their unique expertise and judgement to take a stance on the information uncovered, but also promotes more democratic and objective results. Table 1 presents an overview of the strengths and weaknesses of the critical review.

Table 1: Strengths and Weaknesses of a Critical review - Source: https://www.litr-ex.com/critical-reviews

THE CRITICAL REVIEW	
STRENGTHS	**WEAKNESSES**
Methodologically adaptable: scholars are able to deepen their understanding of complicated issues by evaluating theories and evidence from diverse sources, instead of strictly adhering to one discipline.	Limited methodological guidance.
It gives the possibility to rethink how to interpret a problem.	The flexible boundaries can lead to frustration, as others may not use the same literature or search strategy.
Researchers serve as research instruments, they evaluate and interpret literature using their perspectives, rather than summarising it.	Not suitable for those seeking a definitive or final solution to a specific problem.
Valuable for problems needing innovative thinking.	Hard to delegate this type of review to someone with instructions to follow a specific process.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Using this methodology allowed us to focus on presenting, analysing, and synthesising material from diverse sources. The analysis extends beyond simply describing the identified articles; it includes a degree of analysis and conceptual innovation, rooted in the research team's extensive knowledge of the heritage digitization domain under investigation.

3.2 Categorisation of thematic areas and keywords

REEVALUATE investigates digitisation management with regard to two types of cultural heritage:

1. Cultural heritage being digitised: CH artefacts which are physical and have been digitised afterwards;
2. Born digital cultural heritage: CH artefacts which were created digitally through digital toolset.

To initiate the critical literature review process, we crafted a guideline outlining our methodological approach through collaborative discussions among partners (EFHA, KU Leuven and Arthurs Legal). These discussions aimed to pinpoint and refine specific keywords and thematic areas relevant to our subject matter. The main research question was the following:

> *What are the current ingredients in digitisation management that impact (enable/ inhibit/ create risks) collaboration, creative reuse boosts and promotion of democratic and inclusive prioritisation and contextualisation of digital cultural heritage?*

Once the initial list of these thematic areas was finalised, a research question for each one of the areas was compiled, in order to guide the research on the specific area. The thematic areas were then split amongst the three partners, according to their expertise and relevance - and each partner focused on conducting their part of the literature searches and review. Table 2 presents the list of the pre-defined thematic areas and the corresponding specific research questions selected during the initial stage of the literature review process.

Table 2: Overview of initial key thematic areas

THEMATIC AREA	RESEARCH QUESTIONS	PARTNER
Access and engagement		EFHA
Community involvement, participatory models and stakeholders' needs	Are these relevant elements of the decision-making processes? Do institutions include stakeholders' voices when deciding on digitization priorities, content selection, etc.? Do heritage organisations allow for active engagement from various communities in digitization processes regarding digitised content? Do they employ crowdsourcing in their actions? Do they engage communities in indexing, enriching data, etc.? Is community empowerment a relevant component of institutional policies? Do they employ co-creation activities or collaborative practices in the process?	
Access, use, and reuse of data	Do they impact digitization workflows and content selection? How relevant is the option to open up digitised collections and allow for their reuse?	

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Digital innovation linked to digitization management	How do organisations define innovation regarding digitization management? Is it about the use of technology, innovative workflows, engagement, etc.?	
Digital skills required for digitization	What skills are crucial when it comes to digitization processes? Is reskilling a phenomenon they observe?	
Legal, ethical and policy considerations		ARTHUR
Policy efforts in supporting and maintaining cultural heritage	What EU policies are specifically proposed and implemented to support and maintain cultural heritage in the Union, and more specifically, digitise existing and future cultural heritage?	
Intellectual property rights (Copyrights) and digitization	What is the effect of intellectual property rights, in various forms, on the digitization of cultural heritage? What policies and methods are considered for the intellectual property/copyright/licence tagging, attribution, and management? How is the validation and enforcement of the copyrights/contracts achieved, e.g., preventing out-of-context use or misuse?	
Licensing, Creative Commons, and T&C	What is the effect of various measures, such as contractual terms, on the digitization of cultural heritage?	
Policy and regulatory updates in access, use, and reuse of CH data	How do various EU legislations and policies affect the digitization of cultural heritage (Open Data Directive, Data Governance Act, Data Act, AI Act, etc.)?	
Gaps in EU policies and legislations	How do various applicable policies and legislations concerning intellectual property rights, data sharing, licensing, and the like, affect the digitization of cultural heritage in the European Union? Which gaps in the legal framework present difficulties or opportunities?	
Technology		KU LEUVEN
Data governance/stewardship	What data governance/stewardship policies/methods/plans are proposed and implemented to support the content and metadata management, preservation, provenance, processing, quality, and versioning of (multilingual) data related to cultural heritage? Is there a data management plan?	
Prioritisation	How is it decided which cultural heritage data is prioritised for digitization?	
FAIRness	How is it verified that the digitization and management of cultural heritage data are free of bias, inaccuracies, and stereotypes and promote inclusion and diversity, as well as ethical, sensitive, and respectful (re)use of metadata and content? How is the accessibility, discoverability, findability, reusability, interpretability, interoperability, and trustworthiness of metadata and content achieved?	

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Ownership/Privacy/Legal	What policies and methods are employed to promote the legal access and (re)use of data and prevent the illegal (re)use and alteration of metadata and content related to cultural heritage data? (Authorization, authentication, security, logging) Which measures are taken related to sensitive/personal data?	
Artificial Intelligence	Which AI techniques/algorithms are used to prioritise, contextualise, process, recommend, and manage the content and metadata related to cultural heritage data? Are the employed AI techniques explainable?	
Standards	Which standards/guidelines/best practices are considered for the contextualization, semantic annotation, quality assessment, and sharing/exchange of metadata and content of cultural heritage data?	

3.2.1 Consultation with external experts

After the identification of the aforementioned thematic areas, and in order to validate them and verify the completeness of the thematic areas (i.e., whether there are other areas that are missing from the list), we organised a feedback loop with external experts. An online focus group was conducted, wIth three policy experts, namely a policy officer from DG Connect and two Member States representative in the Common European Dataspace for Cultural Heritage Expert Group (CEDCHE) from France and Poland, allowing us to identify any possible missing topics. This also enabled us to juxtapose and integrate information on the areas already researched.

The received feedback led to the creation of the following additional thematic areas that were added to the ongoing literature search. Table 3 presents these findings.

Table 3: Overview of additional thematic areas

THEMATIC AREA	RESEARCH QUESTIONS	PARTNER
Sustainability		EFHA
Funding	How is digitization financed? What sustainable models are introduced in the digitization management processes?	
Digital sobriety	What are the primary goals of digitization projects, and are they essential? Are we digitising materials that truly need to be digitised, or are we duplicating efforts? What is the environmental footprint of CHIs current digital practices, including energy consumption and carbon emissions? How can the environmental impact of digitization and digital preservation activities be reduced? Are resources (time, money, technology) allocated effectively and efficiently? How can one optimise digitization processes to minimise waste and redundancy?	

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

3.2.2 Final macro thematic areas

At the final stage of the process a total of six macro areas of analysis have been identified: access and engagement, technology, legal and policy considerations, ethical considerations, skills and competences and sustainability. These macro areas are presented in Figure 1.

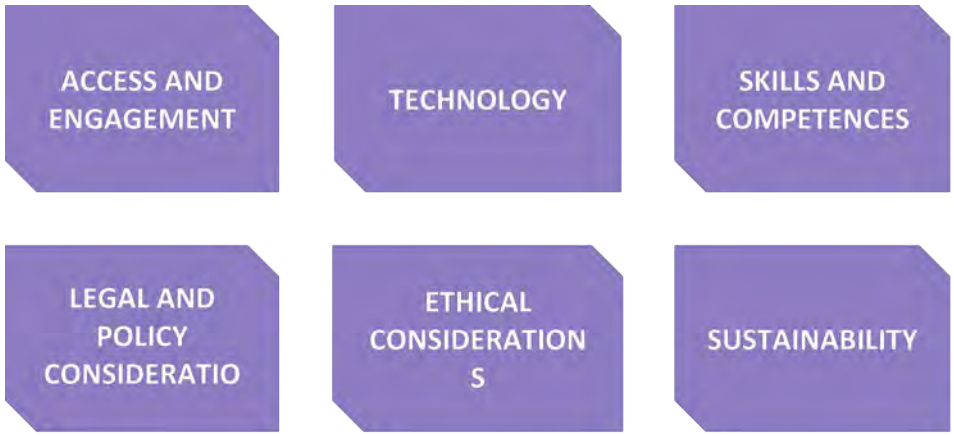

Figure 1: Presentation of the six key macro areas of analysis

3.3 Selection of sources and tools used in the analysis

In terms of categories of sources that act as the pool of information for the literature review, the three key sources of data collection identified are presented in Figure 2.

The data collection aimed at identifying relevant sources following the thematic areas described above. It began by drawing from a pool of relevant articles and projects familiar to the partners. We compiled a list of pre-selected documents from known projects, relying on individual partners' knowledge. Additionally, we sourced further resources from platforms such as Web of Science and Google Scholar, filtering for articles specifically focusing on the European context using our predefined keywords.

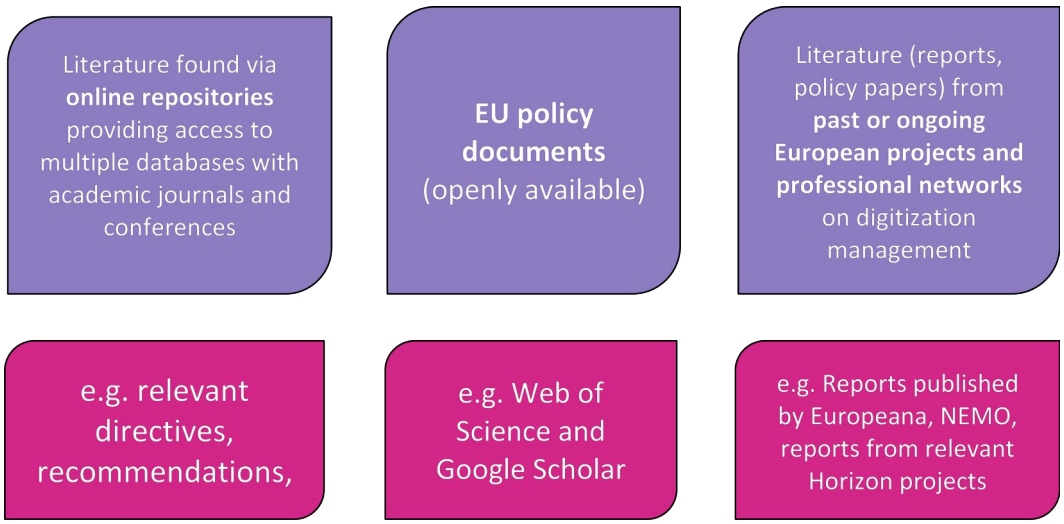

Figure 2: Key data collection methods and exemplary sources

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Moreover, we systematically reviewed EU-level policy documents by leveraging various institutional repositories. The reason for using two different online platforms—Google Scholar and Web of Science—was to broaden the research results and overcome the access limitations faced by non-academic partners involved in the study. Both platforms present similar results when conducting the same research, ensuring a comprehensive and inclusive selection of resources.

The analysis of the extracted literature was assisted by Atlas.Ti1 - an AI-supported software for literature review. The software was used on uploaded articles, employing both manual and automatic coding and categorization techniques to isolate relevant sections corresponding to the predefined keywords. Advanced AI tools were utilised to aid in identifying and extracting pertinent information from the selected papers.

4 Discussion of the key findings

This document provides an analysis from a pan-European perspective, drawing on various sources including EU strategic policy documents, regulations, reports, guidelines from European professional networks, and EU-supported cross-national research initiatives addressing the theme of digitization management.

Despite the clear recommendations from the European Commision (2021)[2] urging Member States to develop or maintain comprehensive and forward-looking digital strategy for cultural heritage to accelerate the sector's digital transformation, many Member States still do not have such strategies or are in the process of creating them. This lack of strategic planning at the national or regional level is one of the main gaps identified in this study, significantly impacting the entire heritage sector.

In countries where digitization strategies for cultural heritage are already in place, these strategies are often integrated into their broader digital cultural policies. State involvement is crucial for financing technical digitization, developing technical standards, aggregating databases, and making digital heritage accessible (Hylland & Primorac, 2024). Therefore, it is highly important for Member States to build their own tailored digital strategy, guided by EU guidelines and objectives.

The document segments the key findings into six building blocks relevant to digitization management: access and engagement, technology, skills and competences, legal issues, ethics and sustainability. The following sections provide an overview of the information discovered during the process.

4.1 Access and engagement

This section presents findings relevant to the accessibility and openness of the digitised and digital CH artefacts to the wider public. Specifically, it discusses the types and levels of involvement of the public to digital CH, the tools for engaging a wider audience, as well as the change required within the CH institutions (skills and processes) in order to accommodate an active and impactful engagement. Moreover, the results stress the need for democratisation of aspects of CH through digital technology, such as contextualization and prioritisation of CH artefact. This democratisation can lead to higher levels of accessibility, increase the relevance of CH to more audiences and ensure a more effective resource allocation, increase transparency, enhance overall quality and build trust. The widening of access can break down geographical barriers, enable and foster reuse, and have a critical social impact.

[1] Atlas.ti, https://atlasti.com/

[2] EC Recommendations on a common European Data Space for Cultural Heritage, https://ec.europa.eu/newsroom/dae/redirection/document/80911

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Lastly, the chapter also presents the challenges and obstacles related to the digitisation of CH artefacts. Namely it describes multidimensional investment that is required, such as financial resources, complexity of metadata standardisation, management processes and technological equipment.

4.1.1 Relevance of participatory practices in CHIs' daily operations

In the digital heritage sector, the value of participatory practices has recently increased significantly, transforming the way cultural institutions engage with the public. This shift in approach has been also reflected and carefully analysed by a number of EU funded projects, like: inDICEs - Measuring the impact of DIgital Culture (Horizon 2020)3, RECHARGE - Resilient European Cultural Heritage as Resource for Growth and Engagement4 or GLAMMONS - Resilient, sustainable and participatory practices: Towards the GLAMs of the commons5 (both Horizon Europe). By incorporating participatory practices, museums, archives, and other heritage organisations can foster deeper connections with diverse communities, encouraging active involvement in the preservation and interpretation of cultural assets. (Drabczyk, Janus, Tarkowski, Ciesielska, Gliściński, 2023). Digital platforms and social media enable broader and more inclusive participation, allowing individuals from around the world to contribute to and engage with heritage collections, thus having a greater impact on society, innovation, wellbeing, etc. (Sacco, 2011). This collaborative approach enhances the richness and relevance of cultural narratives by incorporating multiple perspectives and experiences. A digitisation selection strategy based on stakeholders' needs increases the chances for digital collections to be relevant, widely used, and meaningful. Moreover, it enhances resource allocation, fosters community engagement, enriches digital content, and builds institutional trust, ultimately contributing to the sustainable and inclusive preservation of cultural heritage. Additionally, participatory practices can drive innovation in digital heritage projects, as crowdsourced knowledge and user-generated content can enrich collections, improve metadata accuracy, and uncover previously overlooked artefacts or stories. However, crowdsourcing doesn't immediately mean democratisation of cultural heritage practices (Bonacchi, Bevan, Keinan-Schoonbaert, Pett, Wexler, 2019). Nevertheless, over the years, storytelling and participatory and crowdsourced approaches, based on co-creation of content and collaborations, have become a valuable tool of communication with interested stakeholders (Van Dijk, 2011; Giaccardi, 2012).

The increased value of these practices lies in their ability to democratise heritage, making it more accessible, interactive, and reflective of the collective memory and identity of diverse audiences. This not only strengthens community ties but also ensures that heritage preservation and interpretation remain dynamic and inclusive - creating space for voices that are sometimes omitted or neglected. Therefore, the access discourse requires a distinction between (1) digital access, understood as technological access to the Internet; and (2) digital access, understood as content-related access to collections of books, art, and cultural and historical objects (Gran, Røssaak and Kampen Kristensen, 2019, p. 5). Both areas of access are crucial from the perspective of an inclusive democracy and a sense of fairness.

Nevertheless, introducing participatory practices in heritage organisations requires specific mindset, organisational changes in the institution's structure, and a set of soft skills within the team. These skills will enable the fostering and maintenance of new contributions and perspectives from multiple actors, with their needs and desires recognised as key drivers for the organisation. This represents a significant shift for an organisation traditionally guided by curatorial practices and a custodian-like mindset (Drabczyk et al.,

[3] inDICEs, https://indices-culture.eu/
[4] RECHARGE, https://recharge-culture.eu/
[5] GLAMMONS, https://glammons.eu/

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

2023). Additionally, many institutions operate in a semi-participatory mode, implementing a limited participatory strategy (e.g., inviting audiences to contribute partially to projects designed by the institutions) rather than engaging in extensive co-design exercises and initiatives (Avram, Ciolfi, and Maye, 2020). Therefore, it is crucial to understand the difference in participatory practices and level of engagement, particularly whether participants are involved in decision-making processes and whether the process is transparent (Butterfoss, 2006).

4.1.2 Collaboration with stakeholders

By aligning digitization priorities with the interests and requirements of various stakeholders—such as researchers, educators, local communities, and the general public—heritage institutions can enhance the accessibility and utility of their collections. This approach fosters greater engagement and satisfaction among users, as they are more likely to find the digitised materials relevant to their needs and interests (Drabczyk et al, 2023).

A stakeholder-focused strategy can help identify which collections are most valuable and in demand, ensuring that resources are allocated efficiently and that the most pertinent items are digitised first. This relevance-driven selection increases the likelihood of digital collections being used, studied, and appreciated, thereby maximising the return on investment for digitization projects.

By considering stakeholders' needs, heritage organisations can prioritise the digitization of materials that have significant cultural, educational, or historical importance to specific communities. This can foster a sense of ownership and connection among those communities, encouraging their active participation in the preservation and interpretation of their heritage. Involving stakeholders in the digitization process can help uncover new insights and perspectives that may otherwise be overlooked. Stakeholders can provide valuable input on the context, significance, and interpretation of items, enriching the metadata and enhancing the overall quality of the digital collection (for example via crowdsourcing) and supporting the construction of infrastructures that root them in local practices and needs (Dindler, 2014).

Moreover, a stakeholder-driven digitization strategy promotes transparency and accountability, demonstrating that heritage organisations are responsive to the needs and interests of their audiences. This can build trust and support for future projects, facilitating ongoing collaboration and engagement.

Finally, technology - digital ecosystems, big data and online platforms - place cultural heritage institutions in more complex stakeholder networks that require new managerial approaches. To make their collections visible online and to create economic and cultural value, cultural organisations need to experiment with digital platforms (Pesce, Neirotti, Paolucci, 2019). This requires openness to collaborations with technology suppliers and digital creators.

4.1.3 Access, use and reuse of digital collections

The value of access and reuse of digital collections in the heritage sector offers numerous benefits that extend beyond the traditional boundaries of physical archives and museums. Open digital collections enhance accessibility by breaking down geographical barriers, allowing global access to cultural and historical artefacts. This democratises information and access to knowledge, making it inclusive and available to a wider audience, including those who may not have the means to visit physical collections (Vézina, Benedict, 2024, Wallace, 2021).

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Digitised collections serve as valuable resources for educators and students, providing primary sources for research, study, and teaching. They can be integrated into digital classrooms and learning platforms, enriching educational content. Furthermore, scholars and researchers benefit from easier access to collections digitally, facilitating new insights and discoveries. Additionally, the ability to reuse digital collections in various contexts, such as data mining and AI applications, can lead to innovative approaches in research and analysis. Reusing digital collections can help engage the public through interactive and participatory initiatives. Crowdsourcing projects, virtual exhibitions, and social media initiatives can draw public interest and involvement, enriching the collective understanding of heritage.

However, several challenges and obstacles must be addressed to fully realise these benefits. While digital collections are theoretically accessible to all, disparities in internet access and digital literacy can prevent equitable access. This digital divide (Charles, Willans, Frank, Luz, 2024) limits the reach of digital collections to underrepresented and underserved communities. The unequal distribution of digital resources and access has significant social and cultural consequences. At the same time access to digital technologies has become crucial for individuals to fully participate in modern society (ibid).

Digitization projects, especially those aimed at openness, require significant financial investment and resources. Many heritage institutions, particularly smaller ones, struggle to secure the necessary funding for digitization, digital preservation, and ongoing maintenance of digital collections. Navigating copyright laws and intellectual property rights can be complex, especially when dealing with modern works or those with unclear or complicated ownership, or when new barriers arise from IPR claims on reproduction media or accessible format copies of public domain works (Wallace, 2021). Hence, new and more streamlined IPR management is required.

Effective access and reuse of digital collections depend on comprehensive and standardised metadata. Inconsistent or incomplete metadata can hinder the discoverability and usability of digital assets. Protecting sensitive information and ensuring the security of digital collections are critical. This includes safeguarding against cyber threats and respecting privacy considerations, especially for collections involving personal data. Building a data biography, a comprehensive background of the conception, birth and life of any dataset (Krause, 2019) is essential to secure trustworthy access to digital collections and data.

4.2 Technology

In this section we review the technological aspects related to the digitization in the Cultural Heritage domain, in the context of the REEVALUATE project. To do this, we analysed the web resources, technical documents and publications related to well-known platforms and initiatives in the domain that focus on managing metadata of digitisation efforts. We gather an initial list of well-known initiatives: EUROPEANA[6], the Common Language Resources and Technology Infrastructure[7] (CLARIN), and the Digital Research Infrastructure for the Arts and Humanities[8] (DARIAH). For each platform, we investigated what they offer in terms of artefacts' prioritisation, digitisation and contextualisation, as well as in terms of standardisation, collaboration, sharing and reuse.

4.2.1 Overview of the platforms

Each platform offers a list of related publications, e.g.,

[6] Europeana, https://www.europeana.eu/, last accessed 10 June 2024
[7] CLARIN, https://www.clarin.eu/, last accessed 10 June 2024
[8] DARIAH, https://www.dariah.eu/, last accessed 10 June 2024

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

- Europeana https://pro.europeana.eu/page/publications,
- CLARIN https://www.zotero.org/groups/562080/clarin/library,
- DARIAH https://shs.hal.science/DARIAH and https://www.dariah.eu/about/publications/.

The Europeana platform was launched in 20089 and since then it provides interested parties with access to Europe's digital cultural heritage. It provides access to more than 55 million digital cultural heritage artefacts from more than 3000 institutions across Europe, enhancing their discoverability. Europeana does not store the digitised artefacts or a copy of them in its own platform, but it links to the artefacts from the collections of the institutions relying on aggregators, i.e., organisations which gather data and make it accessible through Europeana10. These aggregators11 consist of more than 40 institutions including Domain & Thematic aggregators, such as Europeana Sounds and European Film Gateway, as well as National & Regional aggregators such as Erfgoedplus.be and the German Digital Library. Since September 2022, the Europeana Foundation is the operator of the common European data space for cultural heritage12, the European Union flagship initiative to accelerate the digital transformation of the cultural heritage sector.

CLARIN was launched in 201213 and is a platform which provides access to data and tools to support research in the humanities and social sciences, and beyond (De Smedt, de Jong, Maegaard, Fiser, Van Uytvanck, 2018). CLARIN provides access to multimodal digital language data (text, audio, video) and advanced tools to explore, analyse or combine these datasets. CLARIN is primarily focused on language resources, including language data and software tools for the preparation, collection, management, or use of the data related to the humanities and social sciences. Cultural heritage data often include texts, manuscripts, and oral histories that can be analysed using these resources.

CLARIN makes use of national nodes to provide the tools and services (CLARIN, 2022). These national nodes are called centres and are labelled according to the expertise and services they provide. Technical centres, also known as CLARIN B-centres, provide access to resources, services and knowledge, while knowledge centres, or CLARIN K-centres, share their knowledge and expertise on one or more aspects of the language domain. CLARIN C-centres, or Metadata Providing Centres, offer machine readable metadata, allowing service providers to harvest their metadata and making them browsable, searchable and combinable. Additionally, D-centres or Other Technical Centres, offer a meaningful service to CLARIN but are not related to language resources (e.g., SAML provider). Lastly, E-centres or External Service Centres offer relevant services but are not provided by CLARIN members14.

DARIAH (Henrich & Gradl, 2013) is another initiative that was launched as a European Research Infrastructure Consortium (ERIC) in 201415. DARIAH aims to enhance and support digitally-enabled research and teaching in the arts and humanities by providing tools and services to facilitate the research for the arts and humanities, while promoting open science principles. As opposed to Europeana that focuses on promoting the access to digitised artefacts, DARIAH focuses on developing an infrastructure and supporting

[9] https://pro.europeana.eu/post/europeana-initiative-marks-15-years-of-empowering-digital-cultural-heritage, last accessed 10 June 2024
[10] https://pro.europeana.eu/share-your-data/process, last accessed 10 June 2024
[11] https://pro.europeana.eu/page/aggregators, last accessed 11 June 2024
[12] https://www.dataspace-culturalheritage.eu/en
[13] https://www.clarin.eu/content/clarin-nutshell, last accessed 10 June 2024
[14] CLARIN centres, https://www.clarin.eu/content/clarin-centres
[15] https://www.dariah.eu/about/dariah-in-nutshell/

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

researchers in using the tools and services it provides them to build, analyse and interpret digital resources. Similarly, to CLARIN, DARIAH is a connector to existing tools and services, for which they provide a catalogue. A distinction is made between Core and Community services16. The DARIAH Core services are mature services owned by DARIAH-ERIC, enabling the infrastructure to carry out its mission, while Community services are mature services owned by one or more DARIAH partner institutions. The latter usually support local capacity building and scientific instrumentation. Due to the focus on the arts and humanities, many of the tools and services that are made available by DARIAH, are relevant to REEVALUATE. For example, the Vocabs Service provides a suite of vocabulary services that are either generally used (e.g., ISO 639-1 Language Codes) or used in the CH field or a specific sub-field thereof (e.g., Iconclass).

4.2.2 Cultural heritage digitisation lifecycle

In the cultural heritage sector, the methodology for the digitisation of artefacts typically involves: prioritisation of artefacts, annotation and contextualisation, collaboration and reuse. In this section, we discuss what the state is of the existing cultural heritage platforms and what the limitations are in relation to the above-mentioned activities.

4.2.2.1 *Prioritisation*

None of the examined platforms offer any dedicated mechanism to support the automatic prioritisation of the selection of artefacts for digitisation and annotation through technological enablers. It seems that the prioritisation of artefacts that will be digitised remains a responsibility of the institutions. The institutions choose which artefacts will be prioritised based on different criteria, e.g., availability of funding, fragility or rarity of artefacts, condition of the artefact, or public interest, etc.

To enable the discoverability of the digitised artefacts, their annotation and contextualisation is required as well as their unique and persistent identification.

4.2.2.2 *Annotation & contextualisation*

EUROPEANA, CLARIN, and DARIAH strive to provide comprehensive frameworks for semantic annotation and contextualisation using metadata standards, ontologies, and controlled vocabularies. Europeana defines the Europeana Data Model (EDM) (Silva & Terra, 2024), its data model which is available in the form of an OWL ontology17. Its formal specification is defined in the EDM Definition and lists the classes and properties that could be used in Europeana. A practical guideline for providers which desire to map their data to EDM is available as “The EDM Mapping Guidelines”. The aforementioned documents are available in the EDM documentation18. CLARIN developed and uses the Component Metadata Infrastructure (CMDI), a flexible metadata framework to manage and describe metadata for language resources and tools in a detailed, customizable, and interoperable manner. CMDI enables the creation of custom metadata profiles (Windhouwer & Goosen, 2022), ensuring interoperable descriptions of language resources. Last, DARIAH offers among its Core Services, the Vocabs Services19 that provides a suite of controlled vocabularies whose (re)use ensures interoperability.

---

[16] https://www.dariah.eu/tools-services/tools-and-services/

[17] https://www.w3.org/OWL/, last accessed 24 June 2024

[18] https://pro.europeana.eu/page/edm-documentation, last accessed 11 June 2024

[19] https://vocabs.dariah.eu, last accessed 11 June 2024

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

The main value of metadata lies in its ability to facilitate discovery. However, the lack of a unified approach and vocabulary across these platforms as well as other cultural heritage platforms hampers seamless data linking and reuse, as well as the overall interoperability among platforms, as different platforms act as data silos. Mappings between the different ontologies used in the cultural heritage sector, e.g., mapping the Europeana Data Model (EDM) and the Component MetaData Infrastructure (CMDI), as well as other relevant standards, could bridge the gap between these infrastructures, enhancing the shareability and interoperability of data in the cultural heritage domain.

Contextualisation refers to the enrichment of the metadata of a digitised artefact with new information or with the creation of new links between the enriched resources and another data resource. In that respect, all platforms offer automated semantic annotation.

Europeana uses the Semantic Enrichment Framework20 to automatically create links from the metadata to controlled vocabularies, i.e., a link from the EDM to a predefined set of resources. Europeana currently dereferences several vocabularies21 from internationally established initiatives or more specific projects, for instance, Europeana enriches places with Geonames22, concepts and agents with DBPedia23, and organisations with the Europeana Entity Collection24. This process is performed based on textual strings in the metadata, but it could be extended with multimodal approaches. Besides annotating the data, Europeana also offers automatic semantic enrichment (Stiller, Petras, Gäde, Isaac, 2014), where the provided metadata is analysed and linked to controlled vocabularies. At the time of writing, the framework was used to enrich data in Europeana with Places (12M), Concepts (34M), Agents (1M), Time Period (Semium) (11M), Time Period (Entity Collection) (200K+) and Organisations (61M) 25.

The CLARIN tool inventory supplies many Machine Learning tools to work with language resources (e.g. Named Entity Recognition, Keyword Extraction, etc.)26. However, the primary use of these tools is not the automatic annotation of data with relevant metadata, but rather enabling more traditional Natural Language Processing (NLP) techniques, such as Named Entity Recognition and Keyword Extraction, as mentioned before. The approaches of both Europeana and CLARIN could be extended with inspiration from the other.

Lastly, DARIAH makes various resources available among its Community Services, for automatic annotation of text (e.g., LinguA27) and images (e.g., eScriptorium28), based on machine learning techniques.

4.2.2.3 *Crowdsourcing*

Each of the platforms uses crowdsourcing as a technique to support one or more of their initiatives. However, it is not used for prioritisation, rather for annotation of artefacts. The CrowdHeritage Platform[29] which

[20]https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit#heading=h.bz78ahr21sjd, last accessed 11 June 2024
[21] https://docs.google.com/spreadsheets/d/1BoDNolkcp_qfvVShdOZyGcf61XslcwKF2MdGcjgYs20/edit#gid=0
[22] Geonames, http://www.geonames.org/, last accessed 10 June 2024
[23] DBpedia, http://wiki.dbpedia.org/
[24] https://www.europeana.eu/en/collections/organisations
[25]https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit#heading=h.bz78ahr21sjd, last accessed 25 June 2024
[26] https://switchboard.clarin.eu/tools, last accessed 11 June 2024
[27] http://www.italianlp.it/demo/linguistic-annotation-tool/, last accessed 11 June 2024
[28] https://test2.fondue.unige.ch/, last accessed 11 June 2024
[29] https://crowdheritage.eu/en

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

was developed under the Europeana Generic Services project CrowdHeritage[30] and enhanced under the CRAFTED[31] project, enables crowdsourced annotation of CH artefacts through the organisation of campaigns. In addition to its hosted solution, the source code is available on GitHub[32].
A smaller initiative has been created related to CLARIN, namely the CrowLL project[33], which gamifies the process of creating manually annotated corpora for teaching and learning purposes, in multiple languages[34]. Similar to CrowdHeritage, their source code is open source, provided on GitHub[35].
DARIAH has a working group on Combining Language Learning with Crowdsourcing Techniques (D4COLLECT) which is aimed at exploring research and innovation trends in the use of crowdsourcing techniques in the domain of language learning, while at the same time opening paths to crowdsource NLP datasets from language learning activities[36].

4.2.2.4 *Persistent identifiers*

A persistent identifier (PID) is a long-lasting reference to a document, file, web page, or other digital object. The Digital Object Identifier (DOI)37, the Open Researcher and Contributor ID (ORCID)38 or the International Standard Book Number (ISBN)39 are a few well-known persistent identifiers. All platforms promote the use of persistent identifiers and several persistent identifier schemes were used in different cultural heritage platforms.

Europeana encourages the use of persistent identifiers (PID) to uniquely identify the artefacts, as well as different entities that are part of the metadata (e.g. institution, aggregator, etc.), however, PIDs are not mandatory and Europeana does not currently offer a service that provides persistent identifiers.

Similarly, to Europeana, CLARIN and DARIAH promote the use of persistent identifiers and they even offer dedicated services for persistent identifiers. Among the services that DARIAH offers in their catalogue, is a PID service40, which allows for the creation, resolving and editing of persistent identifiers. Persistent identifiers are a part of CLARIN's requirements for B centres, improving the reusability of resources and ensuring preservation.

The importance of persistent identification has long been recognized in cultural heritage. Despite their impact on the discoverability and interoperability as well as preservation and long-term access, a recent survey (Freire, Manguinhas, Isaac, Charle, 2023) showed that only 13% of the Europeana records contain a persistent identifier. Among these, half use the HANDLE41 PID scheme, 35% the Archival Resource Key (ARK)42 PID scheme and 10% the Uniform Resource Name (URN) PID scheme. However, records that use

[30] https://pro.europeana.eu/project/crowd-heritage
[31] https://pro.europeana.eu/project/crafted, last accessed 26 June 2024
[32] https://github.com/ails-lab/crowdheritage, last accessed 26 June 2024
[33] https://crowll.celga.iltec.pt/, last accessed 26 June 2024
[34] https://www.clarin.eu/blog/manually-annotated-corpora-teaching-and-learning-crowll-project, last accessed 26 June 2024
[35] https://github.com/nemanja9884/crowl, last accessed 26 June 2024
[36] https://www.dariah.eu/2022/03/16/meet-the-new-dariah-working-group-on-combining-language-learning-with-crowdsourcing-techniques-d4collect/, last accessed 26 June 2024
[37] DOI, https://www.doi.org/
[38] ORCID, https://orcid.org/
[39] ISBN, https://www.isbn-international.org/
[40] https://gwdg.de/services/research-data-management/persistent-identifier-pid/, last accessed 10 June 2024
[41] HANDLE, http://hdl.handle.net/
[42] ARK, https://arks.org/

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

persistent identifiers might follow a scheme that does not have a persistent policy or they may use invalid namespaces.

To enable the contribution of providers as well as the reuse of the digitised artefacts from end users, all cultural heritage platforms offer authentication and authorization mechanisms, as well as metadata related to licence and Intellectual Property Rights (IPR). These two aspects will be further analysed in the following subsections.

4.2.2.5 *Authentication & Authorization*

Different authentication and authorization mechanisms are provided by different cultural heritage platforms to access their records in the form of Application Programming Interfaces (APIs).

Auth Service is Europeana's Authentication and Authorization service. It can be used as an OpenID Connect (OIDC) Provider43, providing Single Sign On (SSO)44 authentication. Currently, this service is used by Europeana to handle registration and management of their user accounts, to grant access to internal and external services and to manage write-level access to Europeana managed services (e.g., the Europeana API).

One of DARIAH's Core Services is an Authentication and Authorization Infrastructure45, based on Shibboleth (SAML)46 (Blümm & Schmunk, 2016). A user is allowed to use a specific application after authenticating for instance by using their institution's account. Additionally, they have to be authorised by the application or by a central Policy Decision Point (PDP). (Kalman, Tonne, Wannenwetsch, , 2015). CLARIN offers a similar authentication and authorization mechanism47 as part of its federated user management and login system using the SAML protocol (Schwartz & Machulak, 2018).

4.2.2.6 *Licence*

Each platform has its own way to provide licences about the digitised artefacts and their metadata, as well as for user-contributed content.

Europeana publishes the metadata it aggregates under the terms of the Creative Commons Zero "No rights reserved" licence (CC0)48, with agreement from the Europeana data providers. Data provided through community collection is similarly published under the CC0 waiver, while user-contributed content is published under the CC-BY-SA49 licence. The Data Exchange Agreement (DEA)50 is signed by Europeana and its data providers (aggregators or singular data providers if needed) to ensure that Europeana is allowed to publish aforementioned metadata, and a preview of the CH artefact, if provided.

For the data itself in Europeana, the data providers specify the rights information in the edm:rights field of the Europeana Data Model, which can then be reused under the specified conditions. This field is mandat-

[43] https://openid.net/developers/how-connect-works/
[44] https://en.wikipedia.org/wiki/Single_sign-on
[45] , last accessed 11 June 2024
[46] https://www.shibboleth.net/about-us/the-shibboleth-project/
[47] , last accessed 11 June 2024
[48] CC0, https://creativecommons.org/public-domain/cc0/
[49] CC-BY-SA, https://creativecommons.org/licenses/by-sa/4.0/
[50] , last accessed 11 June 2024

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

ory and it supports a series of standardised interoperable rights and reuse information from the rightsstatement.org initiative51.

CLARIN has an extensive licensing framework52, which includes general terms of service (TOS), end-user licences and deposition agreements. The TOS regulates access to CLARIN resources. The end-user licence further specifies the access conditions for an end-user to reuse each language resource. The deposition agreements are signed by resource providers who deposit a resource to a CLARIN centre and specify the conditions on which a CLARIN service can distribute a language resource to end-users (Kelli, Vider, Lindén, 2015). The end-user licences are distinguished into public, academic and restricted. CLARIN's standard agreement templates are based on a classification of all language resources into three categories: resources publicly or openly available (PUB), resources for research or academic use (ACA) and resources restricted to individual use (RES) (Kelli et al., 2015).

4.2.2.7 *API availability & interoperability*

Each cultural heritage platform offers its own custom API, hintering the reuse of the digitised artefacts and the interoperability across platforms, as developers need to be familiar with each API.

Europeana provides an extensive suite of APIs53 to access the available records and entities with their associated data and metadata. Notably, they provide a search, a record and an entity API, providing machines with the possibility to search based on a textual string, gather information about a specific record and gather information about a Linked Open Data (LOD) entity, respectively. Additionally, Europeana provides a SPARQL endpoint to interact with the data in an unstructured manner.

CLARIN provides an API, the Digital Object Gateway (DOG)54, to access available data. A PID or URL is resolved by the DOG, by performing a range of standard operations to explore and process the object. As such allowing an end user to access the available Digital Objects.

DARIAH, or rather their Core Service DARIAH-DE Repository, supplies the DH-crud service55 to access publicly available data and metadata. This service is the core storage service for the DARIAH-DE Repository. DH-crud has multiple endpoints, separated by type of metadata to be retrieved (e.g., technical, administrative) and can be requested in multiple formats.

Lastly, an API is available for a joint effort by DARIAH and CLARIN, the Digital Humanities Course Registry (DHCR)56. This initiative supplies information on the range of teaching activities in the DH field worldwide.

4.2.2.8 *Explainable AI (xAI)*

All platforms offer means to search for digitised artefacts whose search algorithms may rank results based on users' queries or browsing behaviour, ultimately influencing the visibility of the artefacts. Nevertheless,

[51] https://rightsstatements.org/en/
[52] https://www.clarin.eu/content/clarin-licensing-framework, last accessed 11 June 2024
[53] https://europeana.atlassian.net/wiki/external/MGU4MjI4ZjA2MmM0NDg3M2JjODQ2ZTZjYzBhZWNhZTg, last accessed 17 June 2024
[54] https://www.clarin.eu/dog, last accessed 17 June 2024
[55] https://repository.de.dariah.eu/doc/services/submodules/dhcrud/dhcrud-webapp-public/docs/index.html, last accessed 17 June 2024
[56] https://dhcr.clarin-dariah.eu/, last accessed 17 June 2024

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

the ranking of the search results is not explainable. While none of the investigated platforms explicitly specialise in Explainable AI (xAI), the infrastructure they offer can indirectly support research and initiatives related to xAI. The investigated platforms try to promote transparency with respect to the methodologies and algorithms used by the tools and services which are connected with the platforms and encourage the researchers to consider the ethical implications.

4.2.2.9 *Cultural heritage digitization lifecycle moving towards data spaces*

In the last couple of years, the European Commission announced several legal frameworks to facilitate data sharing and reuse, e.g., the Data Governance Act (European Data Governance Act, 2020) and the European Data Strategy (A European strategy for data, 2020), as well as to promote trustworthy AI, e.g., the AI Act (Artificial Intelligence Act | Think Tank | European Parliament, 2023). Recently, several initiatives have emerged with methodologies for creating data spaces that enable data sharing and reuse, e.g., the International Data Spaces (IDS), GAIA-X, Data Mesh etc. These dataspaces were deployed in different domains, e.g., agriculture, finances, healthcare, manufacturing etc.

The idea of establishing data spaces for the cultural heritage sector has emerged as well. However, while current cultural heritage platforms may provide a solid basis for such data spaces, they may still need some adjustments. So far the platforms for cultural heritage focused on data accessibility and discoverability, as well as tools and services sharing, but less on defining concrete data sharing strategies, robust data governance practices, and reusable standards that enhance interoperability as in the case of data spaces.

The last few years the cultural heritage field has also moved in this direction, e.g., building on Europeana as the heart of the CH data space and the development of the SSH Open Marketplace, but many aspects need to be further investigated, e.g., authentication and authorization, licensing of data and services and IP management, as well as the establishment of policies and contracts regarding further reuse of artefacts and their enforcement and a more rigid data governance policy.

4.3 Skills and competences

In the heritage sector, the imperative for new skills has never been more pronounced, driven by the rapid pace of digitization and the burgeoning opportunities for online access and reuse of collections, powered also by various technologies (Finnis, 2020). As digital technologies transform how we preserve, interpret, and share cultural assets, professionals in this field must acquire advanced competencies in digital curation, data management, and interactive media. Growing access to and availability of digital technology enables cultural heritage professionals, who may so far have limited experience with interactive technologies, to create materials for and apply these technologies into their daily operations aiming at reaching the audiences (Maye, Bouchard, Avram, Ciolfi, 2017).

These skills are essential to effectively navigate the challenges posed by digitization, such as ensuring the authenticity and integrity of digital copies, managing vast and diverse data sets, and protecting sensitive information. Additionally, proficiency or at least familiarity with emerging technologies like artificial intelligence (AI) and machine learning is essential. AI can revolutionise collection analysis, automate metadata generation, and facilitate sophisticated user interactions through chatbots and virtual assistants. Moreover, enhanced digital literacy empowers heritage professionals to leverage online platforms, thereby expanding audience reach, facilitating collaborative research, and fostering innovative public engagement strategies. Ultimately, investing in these new skills not only can safeguard our cultural heritage in the digital age but also can unlock unprecedented potential for its dynamic and inclusive dissemination.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Furthermore, professionalisation of teams in the sector positively affects the development in CHIs to undertake a digital transformation which is intended to improve their ability to protect and promote their cultural heritage (Palumbo, Ciasullo, Pellegrini, Caputo, Turco, 2022) and sustains the openness of cultural heritage institutions toward relevant stakeholders and user groups as a strategic proposition to upgrade their organisational attractiveness (Maye et al, 2017).

Those new skills need to be approached and understood from different perspectives: from building personal skills and confidence within a perspective of an organisation, influenced and supported by networks (smaller or bigger, formal or informal settings), which sit within the whole heritage sector. The sector, in its turn, must always be considered in the wider societal context (Finnis, Kendrick, 2020).

Digital management should become an integrated professional field in its own right, similar to other areas of responsibility of a CHI (like director of collection, research, conservation, etc.). It requires an experienced expert to manage the area and set professional goals and standards (Sanderhoff, 2014).

4.4 Legal and policy considerations

From the legal and policy-making perspective, the digitisation of CH is a frequently discussed and currently evolving topic. With the pace at which technology has been advancing day by day, national and international regulators are constantly making efforts to steer this advancement by stimulating, regulating or adjusting the direction in which certain technological advancements should evolve. Either purposely or incidentally, some of these efforts also affect the digitisation of CH. In this section, the digitisation of CH is discussed from the legal and policy-making perspective, considering both the status quo of the European law governing this significant and extensive field, as well as the legislators' efforts, aims and ambitions.

In this discussion, it is pertinent to distinguish between the policy and legal framework concerning digitisation of CH and the rules applicable to the way in which a certain party chooses to digitise a CH asset. Even though this deliverable D1.1 considers some of the available methods of digitising CH and their inherent risks, this section mainly has a method-agnostic approach towards the digitisation of CH. The two main topics discussed in this section are the application of IPRs to digitisation of CH and the effects of EU data sharing legislation which has become applicable over the last decade. Since the concept of *"data"* is broad, it is capable of encompassing digitised CH – the consideration of the applicable legal framework is therefore very relevant in analysing and assessing any policy developments and the regulation of digitisation, or digitised CH assets.

One of the aims of this deliverable D1.1 has been the identification and assessment of any gaps or shortcomings in the digitisation of CH. Therefore, where such gaps or shortcomings have been identified, they have been emphasised and briefly discussed, mainly from the legal and policy perspective. The applicable IP regime has overall been found to be fairly robust and provide a range of mechanisms and exceptions that GLAMs (Galleries, libraries, archives and museums) can rely upon in digitising CH artefacts, including the licensing mechanism. Some of the recently adopted copyright legislation, such as the Digital Single Market (DSM) Directive[57], has aimed to facilitate the digitisation of CH assets by GLAMs. However, certain concerns have been identified as far as the efficiency of the proposed mechanism is concerned - those are discussed below.

[57] https://eur-lex.europa.eu/eli/dir/2019/790/oj

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

4.4.1 Intellectual property rights

IPRs aim to reward creative and inventive efforts. These rights provide an author with an opportunity (which may be ring-fenced by certain public interest interests) to exploit the economic value of his or her works. A *right* usually also indicates a *restriction*, as it limits other actors' control over and exploitation of the economic value of the relevant intellectual property asset (Polcak, 2015). Some of the issues in the digitisation of CH, a tremendous part of which is protectable by IPRs, can be attributed to this key principle. Digitisation of CH is often not executed by the author or creator of the relevant CH asset that is digitised. Therefore, it is necessary to ascertain what IPR principles mainly apply in the CH sector, and how these relate to the topic of CH digitisation.

4.4.1.1 *Copyright*

Within the borders of the European Union, the rules concerning copyright differ from jurisdiction to jurisdiction. All EU countries have signed up to the Berne Convention and there are multiple EU directives that are relevant to the topic of digitisation of CH. Before discussing this topic any further, however, it is important to set out some key principles of EU copyright law. The Berne Convention (on which EU copyright law is based, to a great extent) lays down a very broad spectrum of assets qualifying as works, namely every production in the literary, scientific and artistic domain.[58] The work has to be the author's own intellectual creation and it has to be the result of the author's free and creative choices, as well as not being dictated by its technical function (Case C-5/08: Infopaq International A/S v Danske Dagblades Forening).

One of key functions of copyright is to encourage effort and stimulate creativity, innovation, investment and the production of new content, which also extends to the digital environment.[59] Copyright therefore plays a prominent role in the domain of digitised CH. Within the EU, the following legal acts are directly applicable to this domain, specifically the Database Directive, Infosoc Directive, Orphan Works Directive and DSM Directive. It is also worth noting that certain legal acts (such as the E-Commerce Directive) may also be considered relevant, though they are not the focal point of this discussion.

The DSM Directive and CH institutions

The DSM Directive is relevant for CH institutions and GLAMs in a number of ways. First of all, these institutions, along with research organisations, are exempt from the general copyright regime as far as text and data mining for the purposes of scientific research is concerned (Article 3 DSM Directive). Secondly, Article 14 of the DSM Directive stipulates that any material resulting from an act of reproduction of a work of visual art in the public domain, when the term of protection has expired, cannot be subject to copyright or related rights.

Additionally, and most notably, Article 6 of the directive allows GLAMs to make copies of any works that are permanently in their collections *for preservation purposes only*; the making of copies for any other purpose is not permitted. This is relevant in the context of REEVALUATE, because GLAMs would not be able to rely on this exemption when digitising their content for the purpose of further re-use and dissemination. It's worth noting that general re-use and dissemination of the digitised CH assets are generally the purposes of the project's pilots.

[58] Berne Convention, Article 2(1)
[59] DSM Directive, Recital 2

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

Moreover, Articles 8 through 11 introduce a special regime relevant for CH institutions (or GLAMs) concerning the use of out-of-commerce works. These works are defined as copyright works that are no longer or were never available in commerce. The ambition of the EU is, again, to make culture more accessible and available to all. Member States are required to provide a legal doctrine enabling CHIs to digitise these out-of-commerce works and make them available to a wider audience. The directive offers two solutions:

1. the involvement of Collective Management Organisations (CMOs), being organisations that are appointed by rights holders to manage their copyrights, can issue licensing mechanisms;[60] and
2. a copyright exemption.[61]

These solutions are conditional, however. The copyright exemption only applies when there is no CMO present that would be able to issue the licensing mechanism.[62] Also, lawmakers have incorporated an opt-out mechanism enabling rights holders to exclude their works from this mechanism.[63] This also presents a potential cause for caution or concern in the digitisation of CH, as GLAMs cannot rely on a straight-forward copyright exemption mechanism allowing them to make digital copies of CH assets they hold for any purpose.

IPRs in orphan works

When authors of works cannot be identified and/or consent with the use of their works cannot be obtained, because they cannot be located or identified, the works become *orphan works*. For certain beneficiaries such as publicly accessible libraries, educational establishments, museums, archives, film or audio heritage institutions and public-service broadcasting organisations, as such encompassing the GLAMs, the Orphan's Work Directive (OWD) creates opportunities to use those works in their public-interest missions.[64] In the 2011 Impact Assessment of the European Commission, the *'orphan work problem'* was designated as a substantial challenge in creating digital libraries for CH, since only a few Member States had implemented legislation concerning these works and the approach to orphan works had not been harmonised. Because of this, the Commission presented its ambition to create the Orphan Works Directive in *'A Digital Agenda for Europe'*[65] and subsequently issued its proposal.

Specifically, the OWD applies to published written works, cinematographic or audiovisual works and phonograms that are contained in the collections of GLAMs which are protected by copyright or related rights.[66] If the relevant rights holders cannot be identified, the work is then considered an orphan work.[67] Article 6 of the directive stipulates that Member States shall provide for an exception or limitation to the right of reproduction and the right of making available to the public of these works for the earlier mentioned beneficiaries. The exemptions extend to making the work available to the public within the meaning of Art. 3 of the Infosoc Directive and reproducing the work within the meaning of Art. 2 of the Infosoc Directive. The

[60] DSM Directive, Art. 8(1)
[61] Ibid., Art. 8(2)
[62] Ibid., Art. 8(3)
[63] Ibid., Art. 8(4)
[64] OWD, Art. 1
[65] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Agenda for Europe (COM/2010/0245 final)
[66] OWD, Art. 1(2)
[67] Ibid., Art. 2

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

reproduction may only be for the purposes of digitisation, making available, indexing, cataloguing, preservation or restoration.

However, in the Study on the application of the Orphan Works Directive (2012/28/EU) conducted by Milieu for the European Commission, respondents of the survey indicated that the OWD still had significant weaknesses. The main comment of beneficiaries was that they do not have the resources required to comply with the requirements in order to properly use the system.[68] It appears, that in the OWD, the European legislators treaded carefully not to significantly impact the interests of rights holders. As a result, the mechanism is sparsely relied upon in practice (Matas, 2024) and could be of relatively little use when it comes to any efforts for the digitisation of CH. Therefore, it has been identified as a potential gap in the assessment of policy and legislation on CH digitisation.

AI Act

As far as the recently adopted compromise text of the AI Act is concerned, it is worth noting that this legislative text introduces certain obligations on so-called providers of general-purpose AI models. More specifically, such providers shall be required to adopt policies to make sure that their machine-learning models observe EU copyright laws, and to make publicly available a sufficiently detailed summary about the content used for training of their machine-learning models.[69] The term *"providers of general-purpose AI models"* has not been defined or elaborated upon in any more detail, so it may be interpreted relatively widely, and – in so doing – may encompass REEVALUATE participants who make use of machine-learning technologies within the project, or plan to do so. Such participants should be aware of those requirements and should implement mechanisms to comply with the above requirements.

4.4.2 Data sharing

Whilst IPRs can be seen to place certain limits on digitisation of CH, the European Strategy for Data and the subsequent legislative action can be seen as enablers. Through these legislative proposals, EU policymakers aim to extract as much value as possible from all the data that is processed throughout the EU. *"Data"* basically constitutes packages of information in any form, meaning that CH assets may fall within the definition as well. In this section, a closer look is taken at the new data sharing legislation of the EU, with a specific focus on the Common European Data Space for CH. Furthermore, the impact of the GDPR on the digitisation of CH has also been considered, with further assessment of the GDPR in relation to REEVALUATE has been carried out within T6.4 (Data Management and Governance, Regulatory, Societal, Ethical and Gender Issues), more specifically in D6.3 (Data Management Plan and Regulatory, Social, Gender (GEP) and Ethical Issues Report.R1).

Overall, it has been found that the recent policy and legislative efforts have enabled and facilitated the digitisation of CH, rather than stifled it or introduced any significant barriers. As a result, and despite the critical approach to the assessment of the relevant policy documents and legislation, no immediate shortcomings have been identified in this subsection in relation to the digitisation of CH artefacts as such. However, it is worth noting that challenges nonetheless appear to exist in terms of access and reuse of the digitised content, as considered above.

---

[68] European Commission, Directorate-General for Communications Networks, Content and Technology, McGuinn, J., Sproģe, J., Omersa, E. et al., 2021. Study on the application of the Orphan Works Directive (2012/28/EU) – Final report. Publications Office of the European Union. Page 175

[69] AI Act, Art. 53(1)(c) and (d)

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

4.4.2.1 *The Common European Data Space for Cultural Heritage*

In 2020, the European Commission issued the "*European strategy for data*"[70], addressing the EU's need for an innovative, strong and progressive data economy. The document presents the Commission's efforts in reaping the benefits of the overall growth in data volumes, thus ultimately improving the health and well-being of EU citizens, and having a positive impact on the environment, transparent governance and convenient public services.[71] One of the measures designated to achieve this ambition is the creation of a single European data space. This data space is intended to present a single market for data, where the value of both personal and non-personal data can be securely and easily extracted by companies and individuals alike.[72]

The Commission is currently developing Common European Data Spaces in a number of sectors/domains, CH domain being one of them.[73] A *"data space"* is defined as a domain-specific or cross-domain ecosystem of trusted partners that share data (Dobreva, Stefanov, Ivanova, 2022). Even though data concerning CH may not have as direct impact on the wealth or the well-being of EU citizens, innovation in the field of CH can have significant value for the economy and society. It is envisaged that the reuse and preservation of this data will spur creativity across the EU, subsequently stimulating the creation of 'new' heritage.[74] One exemplary factor in this common European data space for CH is *Europeana*, the flagship initiative set up by the EU. This initiative has created standardised frameworks for the online sharing of digital content and accompanying metadata, as mentioned in chapter 4.2.2.

One of the most significant enablers for the common European data space for CH was the previously discussed DSM Directive. The directive laid the groundwork for a more extensive scope of possibilities for GLAMs to operate within the digital environment. With harmonised exceptions for making preservation copies by GLAMs in the public interest and for text and data mining in scientific research, as well as the proposed framework for use of out-of-commerce works, the DSM Directive is one of the fundamental building blocks in the CH data space, as it has paved the way for the increased usability of the various pieces of data sharing legislation for the CH data space.

4.4.2.2 *Relevant EU legislation on data sharing*

The common European data spaces are further enabled through the introduction of several pieces of subject-specific data sharing legislation. This section outlines the key features and considers the relevance of three of the most significant legislative texts, namely the ODD, DGA and the Data Act. This section considers the merits of these various legislative texts in the context of the EU's data economy, as well as their respective impact on the common European data space for CH.

Starting with the ODD, this directive was specifically referred to in the Commission Recommendation on a common European data space for cultural heritage,[75] which recognised its role in setting down the minimum rules governing reuse and laying down practical arrangements to make it easier to reuse existing documents and materials held by public sector bodies of the Member States. The directive has also set out

[70] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data (COM/2020/66 final)
[71] Ibid., page 1.
[72] Ibid., page 4.
[73] Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage, Recital 18
[74] Commission Recommendation (EU) 2021/1970, op. cit., Recital 18
[75] Ibid.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

to stimulate innovation in products and services by governing and facilitating the re-use of certain categories of research data.[76] Under this directive, public sector bodies are required to lower the barriers for the sharing of these categories of documents and information, introducing the new 'open by design and by default' approach.[77]

The directive also specifically addresses the digitisation of CH. In the first place, the ODD acknowledges the importance of public-private partnerships for the use of cultural collections, whilst at the same time facilitating and accelerating public access to the CH. Whereas a certain period of exclusivity would be prudent for the private partner to recoup his or her investment, this period of an exclusive right to digitise cultural resources should not exceed a period of 10 years.[78] Also, the private partner should give the relevant CH institution full rights after the termination of this period. In addition, the ODD addresses the fact that the CH institutions hold vast amounts of digital public domain material, which can form a base for digital content products and services. As such, the works held by them constitute a potentially significant boost for re-use in sectors such as tourism and research.[79]

The DGA complements the ODD, in the sense that it establishes a framework for the secure sharing of categories of data that are protected on the grounds of commercial confidentiality, statistical confidentiality, protection of intellectual property rights of third parties or protection of personal data.[80] Even though the regulation specifically excludes data held by cultural establishments and educational establishments from its scope, since these data are often protected by intellectual property rights of third parties,[81] the DGA also establishes the framework for data intermediation services providers. A data intermediation service constitutes a service that organises commercial relations concerning a data sharing agreement between data holders or data subjects on the one hand, and data users on the other hand.[82] This framework can have a substantial impact on the sharing of works of CH as well.

Finally, the Data Act is designed to *"enhance the EU's data economy and foster a competitive data market by making data [...] more accessible and usable, encouraging data-driven innovation and increasing data availability."*[83] It complements the DGA by providing legal certainty regarding access to and use of data, while the DGA aims to increase trust in voluntary data-sharing mechanisms. Amongst others, it does so by defining essential requirements regarding interoperability to ensure that data can flow seamlessly between the various data spaces participants to share or jointly process data for, amongst others, the development of new products and services, scientific research or civil society initiatives. These requirements are referred to as "essential" and, as such, they are high-level. Amongst others, they require the relevant data space participants to ensure that "the dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty are sufficiently described, where applicable, in a machine-readable format, to allow the recipient to find, access and use the data," and that "the data structures, data formats, vocabularies, classification schemes, taxonomies and code lists, where available, shall be described in a publicly

[76] ODD, Art. 1
[77] Ibid., Art. 5(2)
[78] Ibid., Recital 49 and Art. 12(3)
[79] Ibid., Recital 65 and Art. 12 (3)
[80] DGA, Art. 3(1)
[81] Ibid., Recital 12
[82] Ibid., Art.2(11)
[83] European Commission. Data Act explained. Available at https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained. Accessed 4 June 2024.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

available and consistent manner."[84] It is worth noting that the Data Act's aims and approach complement the European strategy for common European data space in the CH sector (as discussed in detail above) – also acknowledged by the Commission Staff Working Document on Common European Data Spaces.[85]

4.5 Ethical considerations

Traditionally, one of the key aims of professional ethics has been to pursue explicitly defined social values and norms and discourage inappropriate behaviour by professionals in order to develop public trust in services delivered by certain institutions (De Stexhe, Verstraeten et al., 2000). Naturally, the digitisation of CH artefacts has resulted in certain new ethical challenges, including the impact on the management of online engagement with CH collections, certain privacy issues relating to the protection of individuals' personal data in CH documents, ensuring authenticity, organising access to CH objects, carrying out selection and interpretation of digitised artefacts or those to be digitised (Manzuch, 2017), as well as the use and re-use of digitised CH artefacts by and within the context of emerging technologies, such as artificial intelligence or automated decision-making technology. This section summarises the most prominent ethical challenges.

4.5.1 Cultural sensitivity and representation inclusion in digitisation

The accessibility and availability of digitization tools have encouraged the emergence of alternative, non-professional "community" or "independent" archives, including by indigenous communities who have used them to regain control over their traditional knowledge or CH that has often been excessively commercialised. However, this has encouraged GLAMs to grasp the relationship between CH and the worldviews of the societies and communities that create and practise it (Manzuch, 2017). This has, in turn, created additional ethical issues around the digitization process:

(1) When selecting and interpreting CH, GLAMs sometimes apply a "Western" approach and biases, not recognizing or respecting the cultural work and the necessary context. This may cause harm to communities who feel that they have been wrongfully represented;
(2) In working with the CH, GLAMs may choose to apply metadata schemas that the "Western" world commonly uses to describe and arrange indigenous heritage. Again, this may not necessarily be accurate, reflective or respectful in relation to the CH in question; and
(3) Some of the digitised CH artefacts may be considered/intended to be of limited use or access in the relevant community. By digitising them and including them in a repository or reusing them, this quality and/or their uniqueness may suffer.

Some have also opined that, thus far, most digitisation efforts have originated from Western America or Europe, without seeking any input or views of the relevant local communities concerned. As a result, the resulting digital artefacts may be presented or interpreted from the Western perception, irrespective of the local customs (Thompson, 2017). This presents an ethical challenge as well.

4.5.2 Ease of sharing and manipulation

Due to its very nature, digital content can be easily shared, combined, aggregated and modified online. This often raises concerns about ensuring that the content is authentic. GLAMs are often seen to guarantee the authority and trustworthiness, ensuring authenticity of the assets which they provide access to. To main-

[84] Data Act, Art. 31(1)
[85] European Commission, 2024. Commission Staff Working Document on Common European Data Spaces (SWD(2024) 21 final)

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

tain this unique role, it is therefore imperative that the digitised content which they provide access to is authentic and genuine (Thompson, 2017).

There are also certain privacy concerns in relation to personal data of individuals contained in some CH materials – more specifically, there exists a concern that the digitisation of content containing individuals' details may facilitate the undesired dissemination of such details potentially causing harm to one's public image. For instance, a specific example has been referred to in relation to the digitization by an alternative press of lesbian erotic content from 1984 – 2004 that bears historical significance but could be harmful for living persons exposed in the journal (Thompson, 2017).

In addition, some have opined that in order to be able to fully and properly assess and appreciate a CH asset, it is essential that the viewer has the option of assessing and interpreting it from different angles, in various contexts and using all their senses – wherever possible. When it comes to the outcomes or products of digitisation of certain 3D assets, such as historic sites, complexes and other architectural constructions, the viewer often has no choice but to follow a predetermined, recorded view or path, making viewers unable to choose what they see (Thompson, 2017). As some have argued, "*[t]he past should be fully viewable and up to the viewer and the viewer alone to choose which pieces of it they [wish] to interpret as they encounter an augmented cultural heritage site in the field or the museum*" (Kuester et al., 2013). Complementing this argument, it has been noted that certain architectural CH assets or sites change in time (as a result of natural as well geo-political or current-affair events) which digitisation may not always fully account for, again leaving the viewer with an incomplete or inaccurate impression (Thompson, 2017).

4.5.3 Accessibility and online engagement with the digital content

When it comes to accessibility to and engagement with the digital CH content, two key ethical issues have been identified – one concerning insufficient access, whilst the other revolving around unrestricted and non-moderated access. Both will be discussed in turn.

Accessing digitised CH content requires a range of skills, including certain language, communication, social and digital skills, as well as access to the technology enabling the viewing of the relevant CH content. At the same time, some of that CH content in its original form has originated from and can be seen to be entirely dependent on the people and communities who had historically created the CH. Thus, it would be appropriate to allow such people and communities access to the digitised CH content or to encourage their active participation in it. However, some of those people and communities may not possess the skills and resources required and are not able to access and view the digitised CH content as a result. This shows that the accessibility and inclusion of those communities may pose an ethical challenge (Manzuch, 2017).

An ethical issue also revolves around unrestricted and non-moderated access to digitised CH content. Recent online communication trends (including via Web 2.0 and social networks) have enabled users to participate in shaping, interpreting, and disseminating digitised Digital Heritage (DH) content online. This often presents challenges in relation to the ownership of and authority over the content, as some users may intentionally or unintentionally mislead others as to the origins or ownership of the content, or offer or promote various alternative contexts or interpretations. Equally as importantly, however, GLAMs are often not ready for such active user involvement in CH communication, and generally take a rather paternalistic role in defining what is to be communicated and how (Manzuch, 2017). As a result, there remains an ethical challenge in creating an inclusive and participative forum where parties can communicate openly and efficiently, whilst ensuring that participating GLAMs maintain their authority and trustworthiness.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

4.5.4 New funding models

There are some significant costs involved in the digitisation of CH. In order to generate funds to cover those costs, GLAMs increasingly apply business approaches that have become commonplace in raising funds in the private sector. This raises two ethical issues (Manzuch, 2017):

(1) Selection biases. In selecting content suitable for digitization, GLAMs may find themselves intentionally or unintentionally selecting CH assets that may appeal to potential funders, leaving out other content which may be equally as significant or valuable.
(2) Limitation of access. In pursuit of profit, private funders may require exclusive access to the digitized content for a certain period. Such practices raise concerns especially where the projects in question have been partly funded by public funds.

4.6 Sustainability

Sustainability refers to the principle of long lasting, viable and cost-efficient solutions for the CH digitisation. This section summarises the requirements for a sustainable CH digitisation. Namely, it dives into aspects of a transition to a sustainable model of "greener"/more energy efficient technical solutions that can reduce cost – through energy efficient solutions and prioritisation strategies for the digitised artefacts. It also mentions the need for standardised practices and guidelines for such a transition. Moreover, the section also tackles the necessity of funding as a driving force for CHI to operate and digitise and stresses the importance of financial risk management planning and alternative business models for CHIs. Lastly, it deals with the needs for long--term preservation of digital collections and the importance of sustainable digital storage solutions.

4.6.1 Digital sobriety

While there is not much literature available yet on the topic of digital sobriety in digitization management the topic itself is surely gaining on relevance (see Europeana webinars[86] or follow the recently launched project PACESETTERS). Cultural institutions in Europe such as (mostly publicly funded) museums, are destined to play a critical role, by becoming an example, in this sustainable transition taking into account current global, environmental and socio-economic challenges (NEMO, 2023).

Digital sobriety (term put together by the French Shift Project[87]) focuses on minimising the environmental impact of digital projects. This includes also reducing energy consumption during the digitization process, optimising storage solutions, and ensuring that digital archives are managed sustainably. By implementing energy-efficient practices, heritage institutions can contribute to reducing their carbon footprint, better manage digital waste and promote environmental sustainability and awareness (Julie's Bicycle 2024). There is however still little data, best practices and guidance available in the literature on how to implement such a pro-environmental and sustainable approach in heritage organisations.

Efficient management of digitization projects through digital sobriety needs to be more thoroughly investigated as it can lead to significant cost savings. By prioritising essential digitization activities and avoiding unnecessary digital operations, heritage institutions can allocate their resources more effectively.

Furthermore, digital sobriety encourages heritage institutions to focus on the quality and relevance of their digital collections rather than the quantity. By being selective about what gets digitised, institutions can

[86] https://pro.europeana.eu/index.php/event/digital-sobriety-and-glams-a-new-challenge
[87] https://theshiftproject.org/en/article/deploying-digital-sobriety/

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

ensure that their digital collections are more meaningful, easier to manage, and more useful to researchers and the public.

Implementing sustainable digitization practices can be technically challenging. Heritage institutions may need to invest in new, green technologies, such as energy-efficient scanners, servers or cloud infrastructures, and data management systems (Julie's Bicycle, 2024). These technologies can be expensive and require specialised expertise to implement and maintain. It also requires training and incorporation of new sets of skills.

As the topic is recognised as still quite new, developing and implementing policies and standards that support digital sobriety is an ongoing challenge. Heritage institutions lack guidelines that balance sustainability with the preservation and accessibility of cultural heritage.

4.6.2 Funding

Financial costs of massively digitising cultural heritage resources are vast (Lekakis & Dagouni, 2024) and are often beyond availability to many heritage organisations, especially those of smaller size and limited resources.

Most CHIs in Europe are publicly funded, dependent on the financing from public entities. National, regional, and local governments provide significant financial support to cultural heritage institutions. This includes annual grants, subsidies, and special project funding. Also, the EU offers various funding programs and grants for cultural projects through initiatives like Creative Europe[88], Horizon Europe[89], and the European Regional Development Fund[90]. These programs support preservation, digitization, innovation, and cultural exchange.

To ensure their sustainability and resilience, with regards to digitization, but also to digital transition as such, CHIs should develop comprehensive risk management plans to prepare for financial uncertainties and crises, but also allow them to take on risks allowing them to fulfil their public interest mission (e.g. risks linked to copyright and access) (Drabczyk et al 2023). CHIs ought to create long-term sustainability plans that include financial strategies, resource management, and possible diversification of income streams (sponsorship, grants, public-private partnerships, etc.) (Pelissier, 2021) that will support the fulfilment of their plans in the long-run.

4.6.3 Long-term preservation

As flagged in the EC recommendations on the European common data space for cultural heritage (2021) "it is important that Member States continue their efforts to digitise and digitally preserve cultural heritage assets. (…) This would be particularly important for cultural heritage at risk.'

Ensuring the long-term preservation of digital collections involves addressing technical challenges such as data degradation, format obsolescence, and the need for regular updates and migrations to new platforms.

Adopting digital sobriety principles helps in the long-term preservation of digital collections. By reducing the volume of data and implementing sustainable storage solutions, institutions can enhance the longevity

[88] https://culture.ec.europa.eu/creative-europe
[89] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en
[90] https://ec.europa.eu/regional_policy/funding/erdf_en

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

of their digital assets. This approach ensures that digital collections remain accessible and preserved for future generations.

5 Conclusions

The digitisation of cultural heritage offers immense potential for enhancing public access, engagement, and participation in cultural heritage, fostering new collaborations and creative reuses of digital artefacts. However, this transformation is not without its challenges. The analysis presented in this document underscores the necessity of a holistic approach to digitisation management that addresses key areas such as access and reuse, technology, legal and policy considerations, ethics, skills, and sustainability.

For Cultural Heritage Institutions to fully leverage the opportunities fostered by digitisation, comprehensive frameworks and policies are crucial. These should provide clear methodologies for prioritising, contextualising, and managing collections throughout their lifecycle. Addressing issues related to intellectual property rights, digital preservation, and organisational sustainability is essential for ensuring the longevity, reusability and integrity of digital collections.

Moreover, as CHIs navigate the digital landscape, they must embrace new technologies and adopt inclusive practices that amplify previously neglected voices, promoting equity and democratic values. The development of new skills and mindsets among CHI professionals is imperative to adapt to these changes and fully harness the potential of digital heritage.

While significant challenges remain, including legal complexities, ethical considerations, and the need for sustainable practices, the path forward lies in a collaborative, innovative, and inclusive approach to digitisation. By addressing these challenges head-on, CHIs can unlock the transformative power of digital heritage, enriching European society and fostering a more engaged and informed public.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

References

1. *Artificial intelligence act | Think Tank | European Parliament*. (n.d.). Retrieved 11 June 2024, from https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792
2. Avram, G., Ciolfi L. & Maye, L. (2020). Creating tangible interactions with cultural heritage: lessons learned from a large-scale, long-term co-design project, CoDesign, 16:3, 251-266, DOI: 10.1080/15710882.2019.1596288
3. Berne Convention for the Protection of Literary and Artistic Works
4. Blümm, M., & Schmunk, S. (2016). Digital Research Infrastructures: DARIAH. In S. Münster, M. Pfarr-Harfst, P. Kuroczyński, & M. Ioannides (Eds.), *3D Research Challenges in Cultural Heritage II* (pp. 62–73). Springer International Publishing. https://doi.org/10.1007/978-3-319-47647-6_4
5. Bonacchi, C. & Bevan, A. & Keinan-Schoonbaert, A. & Pett, D. & Wexler, J. (2019). Participation in Heritage Crowdsourcing. Museum Management and Curatorship. 10.1080/09647775.2018.1559080.
6. Butterfoss, F. D. (2006). Process evaluation for community participation. Annual review of public health, 27(1), 323–340.
7. Case C-5/08: Infopaq International A/S v Danske Dagblades Forening
8. Charles, E. & Willans, R. & Frank, E. & Luz, A..(2024). Social and Cultural Consequences of the Digital Divide.
9. *CLARIN: The Infrastructure for Language Resources*. (2022). De Gruyter. https://doi.org/10.1515/9783110767377
10. Commission Recommendation of 10.11.2021 on a common European data space for cultural heritage,Brussels, 10.11.2021, C(2021) 7953 final.
11. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Agenda for Europe (COM/2010/0245 final), (2010), 0245 final
12. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data (COM/2020/66 final), (2020), 66 final
13. De Smedt, K., de Jong, F., Maegaard, B., Fiser, D., & Van Uytvanck, D. (2018). Towards an Open Science Infrastructure for the Digital Humanities: The Case of CLARIN. *DHN*, 139–151. https://ceur-ws.org/Vol-2084/paper11.pdf
14. De Stexhe, G., Verstraeten J., et al., (2000). Matter of Breath: Foundations for Professional Ethics. Leuven: Peeters.
15. Dindler, C. (2014). Designing infrastructures for creative engagement, Digital Creativity, 25:3, 212-223, DOI: 10.1080/14626268.2014.904368
16. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases
17. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce')
18. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society
19. Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works (Orphan Works Directive)

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

Funded by the European Union

20. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC
21. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information
22. Directorate-General for Research and Innovation (European Commission), Brunet, P., De Luca, L., Hyvönen, E., Joffres, A., Plassmeyer, P., Pronk, M., Scopigno, R., & Sonkoly, G. (2022). *Report on a European collaborative cloud for cultural heritage: Ex – ante impact assessment*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/64014
23. Dobreva, M., Stefanov, K., Ivanova, K., (2022). Data Spaces for Cultural Heritage: Insights from GLAM Innovation Labs. From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries.
24. Drabczyk, M., Janus, A., Strycharz, J., & Tarkowski, A. (2023). Deliverable 3.1 - Policy Analysis of Value Chains for CHIs in the Digital Single Market - Summary. Zenodo. https://doi.org/10.5281/zenodo.7500819
25. Drabczyk, M., Janus, A., Tarkowski, A., Ciesielska, Z., & Gliściński, K. (2023). Deliverable 3.6: Policy Brief: Towards community-focused cultural heritage institutions in the digital realm. Zenodo. https://doi.org/10.5281/zenodo.7500839
26. European Commission. Data Act explained. Available at https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained. Accessed 4 June 2024.
27. European Commission, (2020). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data (COM/2020/66 final)
28. European Commission, (2021). Study on the application of the Orphan Works Directive (2012/28/EU)
29. European Commission, (2024). Commission Staff Working Document on Common European Data Spaces (SWD(2024) 21 final)
30. European Commission, Directorate-General for Communications Networks, Content and Technology, McGuinn, J., Sproģe, J., Omersa, E. et al., (2021). Study on the application of the Orphan Works Directive (2012/28/EU) – Final report. Publications Office of the European Union.
31. Ferreboeuf, H., Efoui-Hess, M., Marrauld, L., Lescop, C. (2020). Deploying digital sobriety. The Shift Project.
32. Finnis, J., Kendrick, A. (2020). The Digital TransformationAgenda and GLAMs. A Quick Scan Report for Europeana.
33. Freire, N., Manguinhas, H., Isaac, A., & Charles, V. (2023). Persistent Identifier Usage by Cultural Heritage Institutions: A Study on the Europeana.eu Dataset. In O. Alonso, H. Cousijn, G. Silvello, M. Marrero, C. Teixeira Lopes, & S. Marchesin (Eds.), *Linking Theory and Practice of Digital Libraries* (pp. 341–348). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43849-3_31
34. Giaccardi, E. (ed.). (2012). Heritage and Social Media: Understanding Heritage in a Participatory Culture. London: Routledge.
35. Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. Health information & libraries journal, 26(2), 91-108.
36. Gran, A-B., Røssaak E. & Kampen Kristensen, L.-B. (2019). Digital Infrastructure for Diversity—On Digital Bookshelf and Google Books, The Journal of Arts Management, Law, and Society, 49:3, 171-187, DOI: 10.1080/10632921.2019.1581114

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

37. Henrich, A., & Gradl, T. (2013). DARIAH(-DE): Digital Research Infrastructure for the Arts and Humanities — Concepts and Perspectives. *International Journal of Humanities and Arts Computing*, *7*(supplement), 47–58. https://doi.org/10.3366/ijhac.2013.0059
38. Hylland, O.M and Primorac, J. (2024). Rapids and backwaters: Comparing digital cultural policies (2024). In O.M. Hylland and J. Primorac (Ed.) Digital Transformation and Cultural Policies in Europe (pp. 181-208). Routledge.
39. Julie's Bicycle (2024). Environmental Sustainability in the Digital Age of Culture. Opportunities, impacts and emerging practices. Arts Council Englans.
40. Kalman, T., Tonne, D., & Wannenwetsch, O. (2015). Sustainable Preservation for the Arts and Humanities. *New Review of Information Networking*, *20*, 123–136. https://doi.org/10.1080/13614576.2015.1114831
41. Kelli, A., Vider, K., & Lindén, K. (2015). The Regulatory and Contractual Framework as an Integral Part of the CLARIN Infrastructure. *Selected Papers*, *123*.
42. Krause, H. (2019). An Introduction to the Data Biography. Retrieved from https://weallcount.com/2019/01/21/an-introduction-to-the-data-biography/
43. Kuester F. et al., (2013). Digital Archaeological Landscapes & Replicated Artifacts: Questions of Analytical & Phenomenological Authenticity & Ethical Policies in CyberArchaeology, in Digital Heritage Int'l Cong.
44. Lekakis, S., Dagouni, M. (2024). Pandemic-driven shifts of GLAMs finaces and participatory practices: digital policy and management trends in Europe. GLAMMONS project.
45. Manzuch, Z., (2017). Ethical Issues in digitization of cultural heritage. Journal of Contemporary Archival Studies. Volume 4.
46. Matas, A., (2024). The out-of-commerce works system: a promise to unlock our heritage digitally. CeLISR.
47. MAYE, L., BOUCHARD, D., AVRAM, G. and CIOLFI, L. (2017). Supporting Cultural Heritage Professionals Adopting and Shaping Interactive Technologies in Museums. In: DIS '17 : Proceedings of the 2017 ACM Conference on Designing Interactive Systems. ACM, 221-232.
48. NEMO Working Group Sustainability and Climate Action. (2023) Climate protection in museums. Guidelines.
49. Palumbo, R., Ciasullo, M.V., Pellegrini, M.M., Caputo, A. and Turco, M. (2022), "Locally focused and digitally oriented: examining eco-museums' digitization in a service quality management perspective", The TQM Journal, Vol. 34 No. 3, pp. 398-417. https://doi.org/10.1108/TQM-02-2021-0046
50. Pesce, D., Neirotti, P. & Paolucci, E. (2019). When culture meets digital platforms: value creation and stakeholders' alignment in big data use, Current Issues in Tourism, 22:15, 1883-1903, DOI: 10.1080/13683500.2019.1591354
51. Pelissier, M. (2021). Cultural Commons in the digital ecosystem. Wiley
52. Paulk, M. (2002). Capability Maturity Model for Software. In J. J. Marciniak (Ed.), *Encyclopedia of Software Engineering* (1st ed.). Wiley. https://doi.org/10.1002/0471028959.sof589
53. Polcak R., (2015). Digitisation, Cultural Institutions and Intellectual Property. Masaryk University Journal of Law and Technology. Vol. 9(2).
54. Position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389

55. Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)
56. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)
57. Sacco, P. (2011). Culture 3.0: A new perspective for the EU 2014-2020 structural funds programming, European Expert Network on Culture (EENC).
58. Sanderhoff, M. (2014). This belongs to you. In Sanderhoff, M. (Ed.) Sharing is Caring. Openness and sharing in contemporary museum culture. Statens Museum for Kunst, Copenhagen, pp. 20-131.
59. Schwartz, M., & Machulak, M. (2018). SAML. In M. Schwartz & M. Machulak (Eds.), *Securing the Perimeter: Deploying Identity and Access Management with Free Open Source Software* (pp. 59–103). Apress. https://doi.org/10.1007/978-1-4842-2601-8_3
60. Silva, A. L., & Terra, A. L. (2024). Cultural heritage on the Semantic Web: The Europeana Data Model. *IFLA Journal*, *50*(1), 93–107. https://doi.org/10.1177/03400352231202506
61. Stiller, J., Petras, V., Gäde, M., & Isaac, A. (2014). Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. In M. Ioannides, N. Magnenat-Thalmann, E. Fink, R. Žarnić, A.-Y. Yen, & E. Quak (Eds.), *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection* (pp. 238–247). Springer International Publishing. https://doi.org/10.1007/978-3-319-13695-0_23
62. Tartari, M., Sacco, P. L., Manfredini, F., & Pilati, F. (2022). Deliverable 1.7 Guidelines for the best practices regarding the maximisation of the impact of digitisation of cultural heritage. Zenodo. https://doi.org/10.5281/zenodo.7486639
63. Thompson, E.L., (2017). Legal and Ethical Considerations for Digital Recreations of Cultural Heritage. Chapman Law Review, Vol. 20(1). Pages 153-176.
64. Wallace, A. (2021). Accessibility. Open GLAM. Retrieved from https://openglam.pubpub.org/pub/accessibility
65. Van Dijk, D., (2011). Exploring Heritage in Participatory Culture: The MuseumApp. In J. Trant and D. Bearman (eds). Museums and the Web 2011: Proceedings. Toronto: Archives & Museum Informatics. Published March 31, 2011. Consulted June 9, 2024. http://conference.archimuse.com/mw2011/papers/exploring_heritage_in_participatory_culture
66. Vézina, B., Benedict, C. (2024). Don't be a Dinosaur; or, The Benefits of Open Culture. Creative Commons, accessed
67. Windhouwer, M., & Goosen, T. (2022). Component metadata infrastructure. *Clarin: The Infrastructure for Language Resources*, 191–222.

This project has received funding from the EU Horizon Research & Innovation programme under GA No 101132389