REPORT ON A EUROPEAN COLLABORATIVE CLOUD FOR CULTURAL HERITAGE

EX-ANTE IMPACT ASSESSMENT

PREPARED FOR
EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION

PERE BRUNET

LIVIO DE LUCA

EERO HYVÖNEN

ADELINE JOFFRES

PETER PLASSMEYER

MARTIJN PRONK

ROBERTO SCOPIGNO

GÁBOR SONKOLY

Contents

Executive Summary

"If I had to do it again, I would begin with culture" - this recurring bon mot habitually attributed to Jean Monnet is timelier than ever. In an epoch tormented by hot war, pandemic and natural disasters leading to growing social inequalities and political turmoil characterized by populism and anti-democratic tendencies, it is customary to designate secondary importance to culture. Europe has a long history of these controversial periods. Europe also has a long history of appreciation of its unique, diverse and extraordinarily rich culture, which is preserved, studied and exhibited in its museums, galleries, libraries and archives. These institutions are not only the very bearers of European identity, but also the testimonies of the resilience and the adaptive abilities of European culture, which have helped Europeans through many tormented periods. These institutions – in conjunction with European identity building – are experiencing a crucial paradigm shift, which is characterized by accelerating digitization and by new governance models spurred by co-creative science and participation of a wide range of stakeholders.

The conceptual counterpart of this paradigm shift is the current notion of cultural heritage, integrating the diversity (tangible, artistic, natural, intangible, etc.) and the complementing levels (local, regional, national, European, universal) of the centuries-old European heritage. By virtue of this current notion, cultural heritage

1. provides the perspectives to embrace the green transition by re-interpreting the relationship between cultural and natural heritage;
2. sustains social cohesion by engaging citizens, researchers and experts within heritage communities; and
3. protects and transmits tangible cultural assets.

Thus it aims at a common European collaborative space – which is duly reflected in the Horizon Europe Work Programme (Cluster 2). In other words, this European Cultural Heritage is green (sustainable), innovative (providing employment in CCI) and digital. Contemporary digital cultural heritage is not only technologically prevailing, but also leads the way to a new human centered digital world, in which the European Collaborative Cloud for Cultural Heritage (the Cloud) will play a key role.

The objective of this Report is to lay the groundwork for the preparation of the Cluster 2 WP Calls to support the European Collaborative Cloud for Cultural Heritage for the period of 2023-2025, along three milestones:

1. design and implementation of the basic architecture and governance;
2. set up the steering and coordination entity;
3. develop first real-world uses.

The report has been prepared by independent European experts representing the diversity of current Cultural Heritage, as illustrated in the short biographies at the end of the Report. The experts have held twelve meetings between November 2021 and March 2022, in which they thoroughly discussed the context, the possibilities and the specificities of the future Cloud. They could also rely on the outcomes of a Stakeholder Meeting on 12 November 2021, and of the related survey summarized in the Appendix of the Report. The Report is divided into three Parts:

1. General Motivations;
2. Current Status of the Cultural Heritage Domain;
3. Implementation Plan for the Cloud.

Part 1 sets the ground for the preparation of the Cloud. It provides the necessary definitions and key considerations; it describes the general landscape of EC programs, infrastructures and platforms; it presents the relevant scientific and technological trends and pre-establishes assessment and sustainability criteria. The authors clarify *why* (limited commercial interest, ownership of data, importance of European jurisdiction, etc.) and *how* (inclusive, collaborative, interactive, safe, fidelity-

and equality-based, open-access, etc.) the EU must establish and maintain such a Cloud. They list and critically analyse future benefits, such as resource-savings for cultural institutions, interacting with and interlinking knowledge, and boosting the potential for innovation and research. They present and analyse needs and related services requested by the Cultural Heritage sector, as well as the specific technological challenges due to the fast evolution of data types and storage systems. They suggest solutions for how museums and other Cultural Heritage institutions - lacking a unified framework for long-term access to its digital(ized) data - could be interconnected as an engaged community of users in a progressively built and shared, adaptable and open digital ecosystem, which includes tools and services designed in accordance to their specific needs as well as an easy-to-use interface.

Part 2 presents the current status of the Cultural Heritage domain from the perspective of data; requirements and the potential uses. Because of its great complexity, this domain contains a considerable number of data types and, consequently, none of them is considered exclusive or optimal for all the specific applications. The authors explore how the infrastructure aimed at supporting the wide spectrum of Cultural Heritage applications and data uses should be designed in order to

1. fully support the relevant data types;
2. make the content FAIR (Findable, Accessible, Interoperable, and Re-usable); and
3. integrate the appropriate approaches and technologies needed for interacting with Cultural Heritage data.

To support an effective implementation of the Cloud, they propose relevant technical methodologies and tools as well as the selection and assessment criteria of the planned projects. Accordingly, projects should focus first on tools that meet the biggest demand and require the least budget to develop, second on those that meet the biggest demand but are more expensive, and third on those that serve a smaller use base but are cheaper to develop.

Part 3 provides guidelines for the implementation of the Cloud by presenting

1. its architecture (principal motivations, requirements/principles/structure/methodologies, risk assessment);
2. support for the implementation of tools and services using the Cloud data and resources;
3. basic, native cloud-based and add-on tools and services;
4. governance, business model and sustainability.

The governance guidelines are structured around a set of principal requirements (technical robustness and data security, proper governance schema, scalability, networks, training, economic and technical sustainability, integral independent and long-term assessment), to construct a variable-geometry governance. Business models are suggested according to the needs and motivations of museums and other cultural institutions, which could clearly benefit from the collaborative documentation and semantic enrichment offered by the Cloud. The sustainability of the Cloud is considered and presented from the interrelated technical, human and financial perspectives. Special attention is given to potential environmental impacts.

All along, the Report underlines the collaborative, integrative and European character of the Cloud. The Cloud must be collaborative to equally mobilize representatives of museums/cultural institutions, of digital industries and technological development, as well as researchers, specialist and experts on Cultural Heritage. It must be developed by dynamically taking into account these inter-sectoral aspects. It should also play a determining role in the renewal of SSH and in the development of the multidisciplinary European Heritage Studies. The Cloud must be integrative, in which large and small, central and local, decently and poorly financed museums and cultural institutions can equally find their place. The Cloud must be European in order to ensure solid investment and financing, to balance cultural supremacy and to bridge the gap between centre and periphery, both within Europe and within its nation states. Thus, culture could reclaim its fundamental role in the European integration.

Part 1 - GENERAL MOTIVATIONS FOR A EUROPEAN COLLABORATIVE CLOUD FOR CULTURAL HERITAGE

1.1 Motivations and definitions

In this digital age, cultural institutions find themselves confronted with unprecedented challenges. Digitization processes spur the networking of cultural institutions globally and enable new contemporary forms of collaboration. Still, legal concerns and constraints, and organizational hindrances, impede the broad use of existing commercial collaboration systems by museums and other Cultural Heritage institutions.

In light of the above, it is the overall goal to establish a new and innovative digital collaboration working space that caters to the needs and requirements of museum work and makes it easier for museums to share, connect, and collaborate: a European Collaborative Cloud for Cultural Heritage (the Cloud) and collaboration platform for GLAM and other Cultural Heritage institutions.
The development of the platform will foster exchanges between museums, offering a wide variety of content and media, and addressing issues of topical interest regarding museum work. It will speed up discoveries, facilitate understanding, enrich cultural experiences and promote access to Europe's cultural treasures. The digital collaboration space shall encourage further exhibition and research projects, offer coaching opportunities, and facilitate the transmission of know-how, skills, and best practises. It shall provide the opportunity to set up individual workgroups and project structures to simplify workflows.
With the Cloud, the mission is to support research, and to make it more impactful by removing barriers and, also, by keeping data protected. Especially regarding cross-museum collaboration projects, such as exhibition and research projects between two or more partners, communicating and sharing large data volumes is a concern. For this purpose, a collaboration platform supports communication between scientists and between institutions in a legally secure space and aims to provide secure storage space in the Cloud, and tools that improve cooperation and guarantee security according to European law and standards. Results from temporary research projects could be secured here long term and made widely available (for the re-use of other scholars and institutions). The development of an open data protocol is also an integral part of this mission. The long-term availability of the research results and the related open data access protocol will enhance the degree of sustainability which is often challenged by the temporary nature of research projects.

The cultural sector often lacks the financial resources, staff and/or expertise to pursue digital transformations. Therefore, a very important goal for the envisioned Cloud and collaboration space is to ensure access and participation for a wide number of small and mid-sized museums within Europe that would otherwise not have the resources to do so. Joint research and mutual exhibition planning could be facilitated for those currently not able to implement neither digital tools for own museum work nor substantial collaboration with other institutions. An easy, flexible, safe and secure access to a joint digital space could be a booster for a pan-European museum collaboration. A basis for a shared European Cultural Heritage is laid by working together, sharing knowledge and jointly collecting history and researching Cultural Heritage objects.

To unleash the full potential of The Cloud and collaboration space as a tool kit for museums across Europe, additional actions have to accompany the implementation. Regarding technical access to broadband networks and high-capacity internet connections many small and mid-sized museums are currently still cut off from communication streams; especially those in rural areas. For this reason, a concurrent network improvement is mandatory to ensure the success of the collaboration space in a wide area. European Structural and Investment Funds specified for museums could be used to convey the implementation of the collaboration platform, which would provide for reasonable technical preconditions for many museums.

The overall scope of the Cloud is not limited to museums. It can be a valuable resource for all professionals in cultural heritage: GLAM institutions, scholars, students and creative industries and - indirectly, through the exploitation of the research results - to the European public. The availability of digital data and collaborative instruments for creating, managing and disseminating Cultural Heritage knowledge will be a key factor in the evolution to a digital approach to Social Sciences and Humanities (SSH). Hence, the design of the Cloud will rely on the increased understanding of digital democracy stemming from the added value analysis of the state of the art of digital democracy in Europe in the field of museums and Cultural Heritage institutions of all levels from local to European.

The design of the Cloud will be a very complex effort. The Cultural Heritage domain is characterized by many different types of data (audio-visual documents, written documents and those of intangible heritage - food, festivals, arts and crafts, etc.), many knowledge needs and multiple communities of potential users. It will contain and represent tangible and intangible cultural assets and foster a common European Cultural Heritage collaborative space, in which users' needs are wide and differentiated. It will be addressed to the experts of GLAM and other Cultural Heritage institutions (curators, restorers, museologists, archivists, archaeologists, historians, art historians, educational staff members, university and school professors, researchers at the digital departments of Cultural Heritage institutions etc.). From the perspective of the end-users, the development of the Cloud will be dynamic and it will give the opportunity for Cultural Heritage experts and researchers to share research results with a wider public. In this sense, it is important to minimize a deterministic approach in the development of the Cloud and learn from the history of European museums, which gradually integrated the roles of research, pedagogy, public representation then the function of destinations of mass cultural tourism in the last two centuries.
The Cloud should offer instruments for a wide potential user community with different levels of proficiency with ICT technologies. It should provide customer support for the end-users (empowerment, facilitating the easy access to the tools and services) also by means of training and pilots, to guarantee trust and promote community building. The implementation of an awareness and training program will be a key action to ensure wide endorsement and diffuse knowledge on the potential of this Cloud, especially considering the case of staff in small museums. This program must be developed in collaboration with several institutions (representative of the European Cultural Heritage sector) to perfectly match users' needs.

Supporting the needs of Cultural Heritage institutions and museums is not a new direction for the EC; several activities have been already funded within the framework of specific EC WPs. A synthetic review of projects that are analyzing the needs, expectations and/or current use of digital approaches in cultural institutions is presented in section 1.2, entitled *General Landscape*. The support of the EC is necessary to ensure (1) the success of such a project and (2) its sustainability over time, and (3) to offer a sustainable infrastructure, which is the key to the confidence without which the aimed organizations will not bother to get involved in the training and to use the provided services.

The Cloud is primarily a platform for Cultural Heritage professionals: it should support professionals in sharing data, should provide digital tools enabling them to work cooperatively to facilitate the insight and discovery process, and should enable keeping trace of the discovery process (following Open Science directives).

Why a European action? We need a common effort, not just because it requires a solid investment, but because we need a common platform to share and to give structure to European Cultural Heritage, stimulating a collaboration of peers at the EU level. It is crucial to establish an EU approach to Digital

Heritage and related domains. This shared approach is unavoidable for the European Cultural Heritage sector to manage digitization of its great number of collections and to exploit its digital assets in the best possible way in order to reap the benefits of the digital transition and avoid the pitfalls. Every Cultural Heritage institution in Europe should have a feasible opportunity to fully participate in this process.

Being European means that we deal with the linguistic diversity of Europe – not yet equally represented in current databases. It also reflects efforts to balance cultural supremacy within Europe, which favours big nations/cultures/languages over smaller ones. Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies could bridge the gap between centre and periphery, both within Europe and within nation states (for example, helping to translate metadata or training and teaching materials which are often based on texts produced in big languages, building a common thesaurus). As a result of this initiative, content such as object data could be translated automatically, step by step building a European Cultural Heritage thesaurus and facilitating sharing and working together. For example, the metadata of a research paper in Italian would be easily available in Dutch.

One cannot wait for a provision of this type of services by commercial companies. In the first place because the business value of Cultural Heritage is not attractive to big commercial players. Secondly, Cultural Heritage institutions want to avoid the risk of losing ownership and control of their data, a well-known issue with (US) commercial players.

1.1.1 Key notes for a European Collaborative Cloud for Cultural Heritage

1. The European Collaborative Cloud for Cultural Heritage (the Cloud) and working space is a digital infrastructure that will foster **collaboration** between the cultural, creative and technology sectors, in an inclusive way.
The establishment of a digital Cloud and working space can help museums and other cultural institutions meet these challenges and provide **safe, secure and trustworthy communication and collaboration** between museums and cultural partners.
The Cloud shall be a digital platform that offers a knowledge and database for best practices and approved manuals, communication channels, project management tools and much more.

2. The Cloud and working space offers **open access** for the people and institutions involved in research projects to exchange datasets and data files with collaborative workspace and workgroups. (Please note open access may still require sign-up or log-in.)
The Cloud and working space are not a temporary research project, it is a **permanent infrastructure** to support European Horizon research projects and projects in the future FPs.

3. It must be easy to add and to remove users (with specific rights over the data, specified by means of a common **authentication** management).
The Cloud shall be a purposive infrastructure project that will benefit both smaller facilities with few employees and large museums and other cultural organizations in the long run. It is a key stone in the overall transformation process and this project will determine what is required and needed to revolutionize how museums and other cultural institutions interact with one another.

4. The Cloud can pave the way for resource-saving methods by establishing tools and methods that make travel and **art transport largely unnecessary**. That will require complex software, advanced digitization technology, accurate physical reproduction technologies, and a lot of storage space.

5. The Cloud and working space offers the ability to **interact** directly and **interlink** knowledge.

The design and development of the Cloud and working space will foster exchanges between museums, offer a wide variety of content and media, and address issues of topical interest as regards the work of museums and other cultural institutions. In addition, this process will seek to identify new challenges and potential areas of further development of the system. This research and innovation Collaboration Space shall grant accessibility to cultural heritage through new technologies, help in solving research questions with the aid of digital instruments, and encourage and facilitate transmission of know-how and skills, providing the opportunity to set up individual workgroups and project structures, and constituting a European cultural counterpart to commercially driven cloud services. The Cloud is instrumental in creating a true digital European (professional) cultural heritage ecosystem, a place where everyone meets everyone. The Cloud is a powerful tool for smaller institutions to expand their professional network.

6. We need the data to be stored under **European jurisdiction**.
Connection, secure exchanges, communication, and the long-term storage of data are particularly important. In this area, what is lacking at present is a suitable digital platform which offers solutions to the practical problems inherent in cross-museum collaboration, in particular at a European level. Although good strategies and formats do exist, legal concerns and organizational barriers prevent the widespread use of existing systems. These systems are generally developed and marketed with commercial interests in mind and for the most part fall under the jurisdiction of US data protection and exploitation rights. For this reason, there are legitimate fears among curatorial institutions that they will not be able to fulfill their duty and care with regard to the cultural assets entrusted to them, owing to the loss of control over data resulting from the use of these systems.

7. We need independent, non-commercial development and services, which should be driven by the specific interests and needs of the professional communities that we are endorsing, and not driven by commercial purposes and intentions.
The Cloud will be co-developed with researchers of the interdisciplinary domain of Cultural Heritage and with the collaboration of stakeholders of CCIs & Arts to capitalize on the opportunities of the digital transition, avoiding the pitfalls. It has **potential for innovation** and experimentation both in the Cultural Heritage domain (innovating the research and professional practice by using advanced technologies) and in the technological domain, since specific needs of this highly demanding domain can ignite new research questions at the technology design and developments level.

8. We need **encryption and signatures** working over the data. Data should be protected, ownership management of the data should be guaranteed, and it should be possible to log the access and use of data. Perhaps blockchain technology could be a source of inspiration.

9. We need simple ways of **publishing** work results.
Publishing should be very simple and intuitive (supporting the entire chain: from basic data and digital representation of works of art, to the more complex documentation of the professional activities operated over the Cloud and the publishing of final results). Innovative instruments for publishing should be provided, keeping memory of the entire chain and of selected intermediate results (following current policies for the reproducibility of science, which entails also for many professional activities). Here, assistance of Europeana may be highly effective.

10. We need to cope with **long-term data storage** and preservation.
Preserving art and culture is usually measured in centuries or even millennia. Extension of the life time of digital content and digital instruments must be studied and developed for the Cloud.

11. **Equality** as a priority in design and development

The co-creative and integral approach to the Cloud will overcome the numerous inequalities existing in European Cultural Heritage (regional, sectoral, professional, social, etc.). The design of the Cloud is conceived to both **address existing inequalities** and mitigate potential negative socioeconomic effects of the digital transition, which will allow the Cultural Heritage Sector and the Cultural and Creative Industries and the Arts though the involvement of cultural institutions to lead the way to a human-centred digital world.

1.1.2 Some practical examples of needs expressed by the Cultural Heritage community

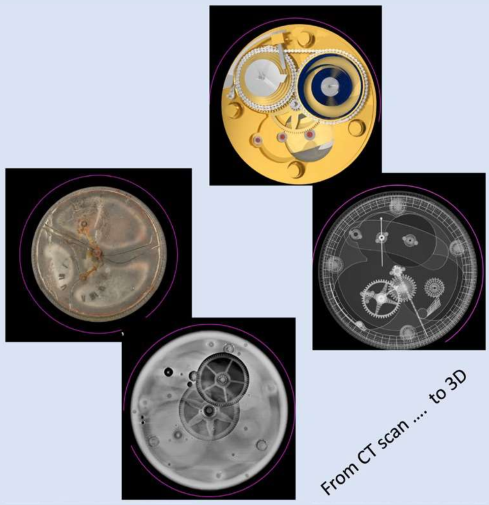
The following boxes are an attempt to make it more concrete, by presenting some examples of Cultural Heritage -related professional activities which require proper support by digital instruments.

Needs - Micro computed tomography scan and animation

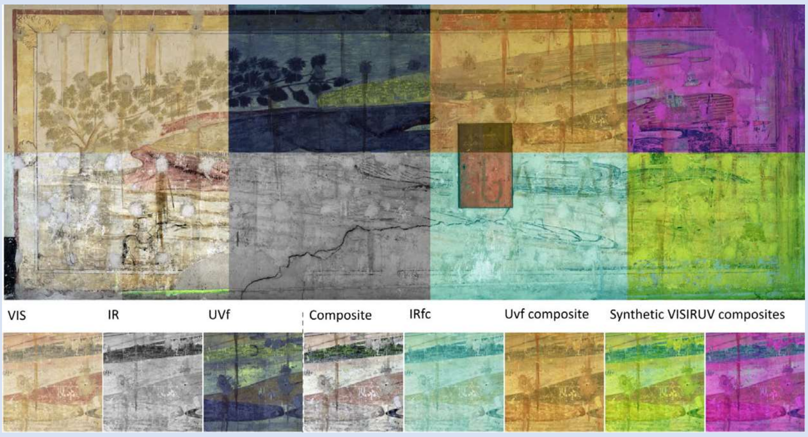
Applied to Pocket Watch Nr. 96, J.H. Seyffert, Dresden 1815.

CT scan can be a very powerful instrument to explore the interior of artwork, to get insight without having to dismantle or destroy them. The study of this ancient clock has shown the potential of CT sampling technologies, supporting the comprehension of the internal mechanisms and the creation of a digital twin.
This experience demonstrated:

- How complex it is to share big data files between people and institutions involved in cooperative research projects;
- How well these technologies allow to compare objects
- from different places without traveling, dismantling and transporting;
- to ensure data storage and security under European jurisdiction.

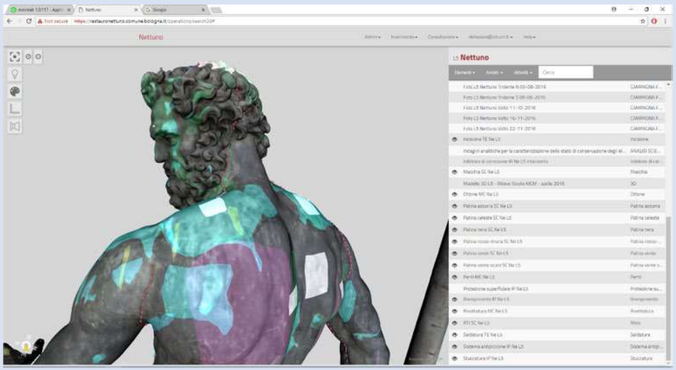


Needs - Assist Cultural Heritage photographers in managing large scientific imagery datasets



Nowadays, photography in cultural heritage documentation is no longer used as a one-off application, but in diverse and complex scenarios involving multiple modalities: sensors, scales, spectral ranges and time periods. The integration of emergent solutions for automatic 3D image spatialization in a dedicated cloud infrastructure should enable the implementation of an iterative image registration process that allows curators, conservators and researchers to manage large and complex scientific image data for the continuous documentation of heritage artefacts. In this context, the joint registration of photo documentation with other digital assets is also an important step towards data fusion and collaborative analysis scenarios with potential impacts for large-scale comparative studies.

Needs – Assist Cultural Heritage Restoration with documentation and analysis instruments



Cultural Heritage restoration is a complex activity, requiring an intense preliminary analysis of the artwork, study of previous interventions, planning of a series of scientific investigations/analysis, documenting the conservation conditions with maps and photographs, planning and executing the intervention and, finally, documenting the work done and the final results obtained. All these activities produce a large corpus of data and documents (text, many different types of visual data, drawings). Curators/restorers/scientists need to: properly store those data on a common digital **repository** accessible to the restoration team; geo-locate the information contained in these documents (since many of them will be related to specific locations of the artwork); finally, they should be able not just to interactively **visualize** and **analyze** those data, bot to correlate and integrate the different data to increase their insight capability. **Data integration/correlation** is key in discovering new knowledge and enabling an improved insight.
Moreover, restoration is a **cooperative work** of a multi-disciplinary team. Thus, a supporting instrument should enable cooperative work sessions, and should allow to insert **annotations** which enable to encode in the data (and preserve) both temporary notes and final results (contributing to). Ensuring **preservation** is a strong need.

Needs - Interactive kiosks for exhibitions and artwork presentation

Designing and preparing kiosks for exhibitions and permanent presentation of digital artwork and monument reproductions, is not a trivial task for museums. The design should be user friendly, as visitors usually possess limited skills in managing 3D interfaces. It should be immune to damage and to the environment, being also accessible by as many people as possible, including people with disabilities.

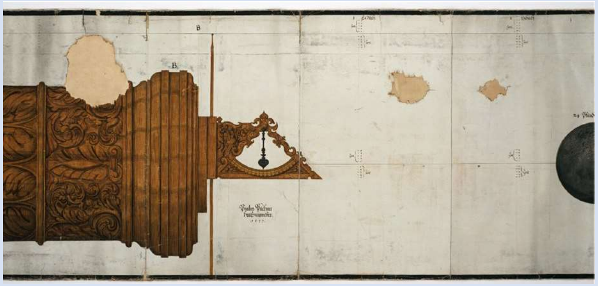
Both the type of presentation (either an interactive inspection or an immersive experience) and the specific design of the user interface should be tailored to the specific characteristics of the objects, monuments or environments to be shown, also taking into account the interests and skills of the targeted visitors. They may consider exploration tasks with no explicit goal other than gathering information on the artworks, searching specific details or places, fine-tuning to precisely examine an object from slightly different angles, or looking into hidden and even unreachable details that are hidden in real monuments. The overall design should facilitate enjoyable and culturally rich user experiences, connecting their visual perception with metadata, historical information, old pictures and narratives.

Kiosk design is not trivial, including many choices. For instance, using new interaction paradigms like virtual reality may not always be the best solution. While being attractive to groups of young users, it may retract other visitors from going to the experience. Thinking on interactive systems allowing one active user and multiple passive viewers can be a good solution, as visitors are not forced to manipulate advanced user interfaces. Also, merging advanced interaction tools at specific places with automatic or video-driven transportation among viewpoints could be considered. On the other hand, mixed-reality schemes could be really useful to enrich visitor's experience at the real sites, by showing different historical narratives in a visual way, for instance.

Museum curators would profit from having flexible tools to design and build these kiosks from existing digital models, also having tools to simulate and compare alternate solutions during the design phase. Reaching a usable design is essential, as insufficient navigation support may cause people to become disoriented or even get lost.

.

Needs - Open access: to read and to write



The “*Shot at noon - European artillery instruments c. 1500-1750*” is a vivid example of an international cooperative project. It involves an international research group (5 scholars in 4 museums in 3 countries) which is studying a corpus composed of 150+ objects in 60+ museums all over Europe, the USA and Canada.

This activity requires instruments and facilities for:

- exchange of datasets and data files
- collaborative workspaces and workgroups
- simple adding and removing of users
- direct interaction possibilities
- data storage under European jurisdiction

1.2 General Landscape: EC programs, research infrastructures, digital platforms, portals

This section presents a list of specific needs and related services requested by the museums and Cultural Heritage community, which should be fulfilled by a European Collaborative Cloud for Cultural Heritage (the Cloud). These services range from basic and general purpose, to more specific and strongly Cultural Heritage -oriented:

- **Data space**: provision of raw data space for storing files;
- **Computing resources**: provision of raw processing resources for running user codes (including specific HW resources, e.g., special purpose architectures for Artificial Intelligence (AI) codes);
- **Virtual Research Environment**: basic instruments supporting the cooperative work of groups of scientists/professionals (common data space, discussion instruments, code execution, ...); these instruments often include **authentication** services;
- **Catalogue**: provides an indexing over the different resources and tools offered by the members of a community or consortium (index of resources developed and managed independently by several partners);
- **Repository**: digital archive of structured data (including provenance, access rights, data quality information, etc), which includes search & retrieval interface, deigned to serve a community (distributed or centralized); here it would be useful to differentiate among actors who ***host and manage*** a repository and actors who provide the technical ***instruments to implement*** a repository; moreover, we should also consider the case of actors which implement ***portals*** (indexing digital materials stored elsewhere):
- **Semantic data**: management of semantic data, data enrichment;
- **Visualization tools**: services supporting the interactive visualization of data (either serving a single or multiple data types);
- **Data Analysis**: services supporting more sophisticated analysis of data (either visually, numerically, AI-based), to support insight or to enhance the discovery process (e.g., AI-based data analysis solutions);
- **Software as a Service**: platforms which provide services for working over the data with SW tools, enabling data transformation, enhancement, process documentation, etc. (should also include a proper interface to add new services on top of the platform);
- **Open Science**: support for management of data (including data concerning publications)
- **Assessment**: instruments to enable the assessment of data and services (real use, impact, effectiveness);
- **Data Preservation**: instruments enabling persistent preservation of data and access to data through durable tools.

These services will be described in detail in Part 2 (community needs and technical constraints) and Part 3 (our vision on how these should be implemented and offered to our community).

This list of services is a basis for analyzing what is now available (thanks to the results of previous/current projects and infrastructure efforts). This is the main scope of this subsection, first to use a common vocabulary while describing synthetically the current status, and then to present a comparative resume in Table 1.

The demand for/project of the creation of the Cloud specifically dedicated to cultural heritage institutions comes in a context where the complexity of the organizational landscape increased substantially, both in terms of objects (data, digital models, etc.) and solutions to preserve them and to give citizens access to this information.

On the one hand, there are public systems, mainly linked to research data (in the vein of programs aimed at encouraging innovation and research), which are more or less closely related to the problems and needs of cultural institutions, and which have grown exponentially since the beginning of the 2000s. Then there are the GAFAM companies (Google, Apple, Facebook, Amazon and Microsoft) which, having grasped the technological and strategic stakes of these issues early on, are now proposing commercial solutions by offering cloud services (storage, web services, etc.), which raises, beyond technical issues, political, legal and ethical questions, which are very sensitive in the field of Cultural Heritage.

These questions are also valid in other, equally sensitive areas, such as health or business. In these areas, national and/or the future European Collaborative Cloud for Cultural Heritage initiatives are emerging (e.g., GAIA-X infrastructure).

In this context, it seems important to us to first distinguish between commercial cloud services (general purpose), or existing initiatives proposing generalist clouds supporting science, which could also be applied to the cultural heritage field and institutions.
Within the framework of the creation of the European Research Area, European infrastructures have also seen the light of day and it is important to take stock of what they bring, on the place they would have alongside the Cloud, if they do not directly meet the needs expressed by the latter, as museums and other cultural institutions are closely linked to research. Most of these different organizations are directly or indirectly included in the governance of the construction of the European Open Science Cloud (EOSC) which is ultimately a federative type entity. This is why we felt it was important to detail the contribution and role of those closest to the Cultural Heritage domain, or which cover it in part. Finally, the landscape is also completed by initiatives, organisations and projects more directly linked to the Cultural Heritage field.

After having described and explained the contributions of these structures, it will be necessary to measure their contributions, in terms of services, innovation and sustainability, to cultural institutions, in order to see how some of them could be articulated with a proposal for the Cloud, if necessary. This is also the scope of Table 1 presented in subsection 1.2.6.

1.2.1 Commercial general purpose cloud services

The first type of support is offered by commercial companies, the well-known cloud services. This is just the basic offer (even if it is a considerable technological advance) and should not be confused with the European Collaborative Cloud for Cultural Heritage proposed in this Report, since those platforms offer only general-purpose instruments.

A quite consolidated terminology defines three different cloud levels for providing services. In extreme synthesis (see also Figure 1 below) users are offered:

- **on site/ no services:** the user has everything he needs on his computer (pre-cloud situation);
- **IaaS (Infrastructure as a Service)**: offers basic infrastructure services, such as data storage, processing resources, virtualization; the user remains responsible of data (structure, encoding) and provides the software instruments working over the data;
- **PaaS (Platform as a Service)**: it adds the support for writing applications (the main user is still a programmer), offer an environment for developing applications;
- **SaaS (Software as a Service)**: the more complete offer of cloud services, includes also one or multiple applications, working over a defined data model, and offered via a web browser.

The model we foresee for the European Collaborative Cloud for Cultural Heritage is a **SaaS** approach (offering applications and data to the final users), where the software services are designed following the needs of the Cultural Heritage communities. It will also offer PaaS support to other future consortia willing to extend the functionalities provided by the European Collaborative Cloud for Cultural Heritage, guaranteeing also its evolutionary character and the cutting edge of the needs and technologies that are evolving very rapidly in the field.

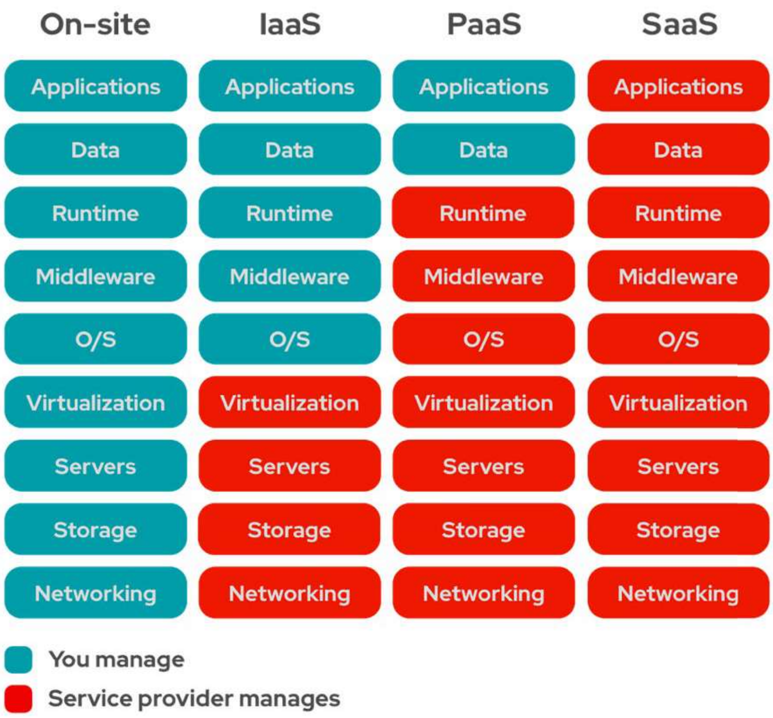


Figure 1: The four levels of cloud services (On-site, Iaas, Paas and Saas), according with the visual presentation proposed by RedHat (image at: https://www.redhat.com/en/topics/cloud-computing/iaas-vs-paas-vs-saas) [1].

While focusing on the group of generalist infrastructures, we should also list and consider private initiatives (proprietary clouds, commercial support tools, etc.) and, on the other side of the spectrum, we will describe existing open-source solutions.

Possible incarnations of **commercial cloud systems** are the facilities offered by IT giants, such as:

- **Amazon**: it offers **Amazon Drive** (basic services for archival of collection of images, video or file) or **AWS** (a complete SaaS cloud, offering a complete catalogue of services);
- **Microsoft**: it offers the **Microsoft Azure** cloud platform, offering IaaS, PaaS and SaaS services. The services provided by Microsoft are efficient but less recommended if one wishes to develop open solutions, as their advanced "in-house" technologies (.net) are favoured. If a choice towards a commercial solution had to be made, it would probably be less coherent to turn to Microsoft for this reason;
- **Google**: form basic collaborative data sharing provided with GDrive, to Google Cloud Platform a complete platform of cloud services, but technically less reliable than Amazon and Microsoft;

- **Dropbox**: it offers data space, a file hosting service a photo uploading service;
- **Mega**: it offers cloud storage and file hosting services.

These proprietary clouds are also paired by some open-source cloud initiatives, including

- OwnCloud (https://owncloud.com/);
- NextCloud (.https://nextcloud.com/);
- Pydio (https://pydio.com/);
- Ceph (https://ceph.com/en/).

1.2.2 Generalist European initiatives focusing on data and enabling scientific collaboration

Here we review the initiatives which support research to comply with Open Science and FAIR data management (https://www.go-fair.org/fair-principles/), providing guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets, and offering general-purpose resources, i.e., not specifically designed to fulfill the needs of the Cultural Heritage sector.

European Open Science Cloud - EOSC (https://eosc-portal.eu/)
EOSC is the more ambitious generalist initiative. It is a project that integrates and federates existing infrastructures of different types, which we will analyze below. According to the project webpage, the EOSC is "an environment for hosting and processing research data to support EU science" (https://eosc-portal.eu/about/eosc). The goal is to "provide European researchers, innovators, companies, and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes" (https://eosc-portal.eu/about/eosc). The EOSC is the operational infrastructure that will enable the implementation of the recommendations for **Open Science** in different domains and enable FAIR Data management practices. The EOSC also aims to provide services to implement this vision.

To meet this goal, different calls have been opened in the framework of the H2020 program, and several domain-oriented clusters have been created. To some extent, the Cultural Heritage field is partially covered by the SSH cluster with the SSHOC project (https://sshopencloud.eu/). The main output of this cluster will be to design the governance of integrated and "seamless" SSH in a European Open Science Cloud, and to tool it, in particular with the creation of the SSH Open Marketplace. In addition, particular attention will be paid to the creation of a continuously trained community through adapted training content and partnerships with national and European infrastructures.

OpenAIRE (https://www.openaire.eu/)
OpenAIRE is an e-infrastructure providing services to support the dynamics of **Open Science** and **FAIR data** production and management. The OpenAIRE project, in the vanguard of the open access and open data movements in Europe was commissioned by the European Commission (EC) to support their nascent Open Data policy by providing a catch-all repository for EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability. OpenAIRE's mission is now closely linked to the mission of the EC: to provide unlimited, barrier free, open access to research outputs financed by public funding in Europe. So, OpenAIRE fulfils the EOSC vision substantially, as its operations already provide the glue for many of the user and research driven functionalities, whether these come from the long tail of science (repositories and local support) or domain-related research communities or Research Infrastructures. OpenAire is actively involved in the implementation of the EOSC, in particular through its portal, which it feeds with updated contents, dissemination activities, and conveys the community feedback on the portal for future improvements.

One of the flagship services from OpenAIRE is the "catch-all" **repository Zenodo** (https://about.zenodo.org/) . It was launched in 2013 and now it is listed as an aggregator service in the EOSC catalogue. In support of its research programs, CERN has developed tools for Big Data management and extended Digital Library capabilities for Open Data. Through Zenodo these Big Science tools could be effectively shared with the long-tail of research.

EGI (https://www.egi.eu/
Another important e-infrastructure interesting to mention in terms of cloud services offered is EGI. EGI is a federation of twenty-one cloud providers and hundreds of data centres, spread across Europe and worldwide. What is more, it is committed to sharing the benefits of e-Infrastructure technology and services with industry, especially SMEs. In terms of the services offered to research, EGI delivers advanced computing services to support scientists, multinational projects and research infrastructures

D4Science (https://www.d4science.org/)
Another data infrastructure that deserves to be mentioned is D4Science. It develops services in the framework of European projects, including in the Cultural Heritage field (e.g., PARTHENOS and ARIADNE(Plus) H2020 projects). D4Science is an organization of the same rank as other infrastructures (data and e-infrastructures) offering a data Infrastructure since 2009. Indeed, D4Science is connecting +15,000 scientists in 50+ countries; integrating data from +50 heterogeneous providers; executing +50,000 data analysis/month; providing access to over a billion quality records in repositories worldwide; operating with 99,8 % service availability.
D4Science hosts +150 **Virtual Research Environments** (VREs) to serve the biological, ecological, environmental, social mining, culture heritage, and statistical communities world-wide.

D4Science offers basic infrastructure resources to support the collaborative work of scientists: the creation of specific VREs to support research groups which include a common data space, the provision and virtualization of computing resources.

1.2.3 European organisations (research infrastructures, initiatives) focusing on Cultural Heritage

European Research Infrastructure Consortia (ERICs) are essentially distributed research infrastructures of European scope. Since the emergence of these initiatives in the early 2000s, ERICs have become embedded in the European research landscape and are included in the ESFRI roadmap (https://www.esfri.eu/esfri-roadmap-2021), to which the national infrastructure roadmaps are generally aligned. In the context of our prospective work, these infrastructures are of interest to us because they aim to improve the knowledge and use of digital tools and services by researchers in the field, according to the ambitions and principles of Open Science.

Examples of such initiatives with a central interest and focus on Cultural Heritage are **DARIAH** (which stands for Digital Research Infrastructure for Arts and Humanities), **CLARIN** (which stands for Common Language Resources and Technology Infrastructure), and more specifically (and recently), **E-RISH** (which stands for European Research Infrastructure for Heritage Science).

DARIAH-EU (https://www.dariah.eu/) is one of the first ERIC initiatives that has started to structure the digital Arts and Humanities landscape at the European level since 2006, when it was included in the first ESFRI roadmap publication. DARIAH-EU aims to enhance and support digital research and education in the arts and humanities. As an ERIC, it operates as a distributed European infrastructure based on national nodes. This articulation aims to give more visibility to national initiatives, tools and services at national and European level, so that research is as open as possible and researchers have a better knowledge of the tools available in their field and can eventually endorse them. Despite this, the consolidation of ERIC's governance and scientific objectives, notably through the Humanities at Scale (http://has.dariah.eu/; 2015-1019) and DESIR projects, has recently produced new perspectives in terms of services to the community, by DARIAH-EU itself.

Thus DARIAH-EU has recently reoriented its activities and strategy around four pillars according to its Strategic Plan 2019-2026:

1. The coordination of the creation of a Marketplace dedicated to Social Sciences and Humanities (SSH) to facilitate fluid exchange of tools, services, data, and knowledge. It is within the framework of the SSHOC project that DARIAH-EU is coordinating the creation of this Marketplace adapted to the Arts and the SSH ((https://marketplace.sshopencloud.eu/), in order to meet the challenge of access to **tools**, **data**, **knowledge** and **services** optimized for researchers in these fields. For example, this Market Place, recently set up after its test phase, uses tools developed within the framework of the H2020 PARTHENOS project, which dealt with cultural heritage in the broad sense of the term and thus integrated the arts and human and social sciences into the scope of DARIAH-EU. This easy entry point is developed as a component of the European Open Science Cloud, as mentioned above.
2. The second pillar focuses on increasing access to **education and training**. In a rapidly changing environment, research requires sustained access to new perspectives and methods. DARIAH-EU promotes awareness and development of skills outside formal qualifications, complementing the formal education provided by our university partners. Resources produced in different member country contexts are aggregated with the "DARIAH Campus" platform (https://campus.dariah.eu/), an outstanding outcome of the H2020 DESIR project (https://www.dariah.eu/activities/projects-and-affiliations/desir/). DARIAH-EU is working to advance this topic through dedicated working groups.
3. The creation of **Working Groups**, Hubs and other forms or more Transnational and Transdisciplinary organization is an organizational novelty that has reinforced the consideration of communities and their know-how and skills in order to stimulate the exchange on these themes in SSH at the European level. DARIAH-EU enables and encourages communities to self-organize around emergent research themes and supports them by providing structure, communication channels, funding, networking, and traveling opportunities.
4. Finally, the fourth pillar is to support the building of bridges between research policies and communities of practice. In this, "DARIAH is an advocate for researchers who create or use digital tools, methods and services in the arts and humanities, and represents their concerns and interests at the European level when research policies are discussed" (https://www.dariah.eu/about/mission-vision/).

Currently, DARIAH-EU has twenty countries as members, one observer and six co-operating countries as partners from where several institutions are participating in DARIAH's activities.

Overall, DARIAH acts more as a connector to existing services, through its organization that mobilizes national coordinators and transversal thematic working groups and with the support of its Open Science Officer; but its role is also to empower European researchers to ensure that they are able to:

- **assess** the impact of technology on their daily research work in an informed way;
- **access** the data, tools, services, knowledge and networks they need seamlessly and in context-rich virtual and human environments;
- produce excellent **digital technology-based scholarship** that is reusable, visible and sustainable, in line with FAIR principles.

Thus, the structure is not directly intended to provide cloud type services, and even less specifically dedicated to cultural institutions, but rather to ensure that researchers are able to understand the issues at stake, so that they can make informed use of them during their research.

CLARIN (https://www.clarin.eu/)
Created in 2012, CLARIN is the pan-European research infrastructure for language as social and cultural data, the latter aspect being of interest to the European Collaborative Cloud for Cultural Heritage in the context of this report. Within CLARIN, the term '**language resources'** refers to a broad range of speech and language data types and resources in machine readable form, and tools and services for the processing of **language data**. Following a longstanding tradition [2], the term 'language resources' also covers **software tools** for the preparation, collection, management, or use of language-related data. Examples of such tools are corpus management and exploration systems, OCR systems, pipelines, speech processing systems, machine translation systems and environments for manual annotation and evaluation.

CLARIN offers access to data, tools, and services to support research based on language resources. CLARIN therefore has a narrower scientific scope and more specific objectives than DARIAH. Also, even though the structure based on national nodes remains the same, its organization and the type of service it provides is different.

Indeed, CLARIN is specific in that it certifies National Centres according to the skills, tools and services that these centres are able to provide to local communities in member countries. According to a protocol and established criteria, centres are labelled "B-Centres", of national relevance, providing both resources and centres, "K-Centres" (Knowledge Centre), "C-Centres" (Metadata Providing Centres, also beyond the limit of CLARIN membership). The CLARIN ERIC also encourages **training**, by offering support for dedicated technical training. It also promotes scientific exchanges by financing short stays in CLARIN centres to promote the use of its services and to train researchers.
The tools put forward by the infrastructure are mainly those of the CLARIN centres (**depositing** services and **long-term preservation** for example). However, the infrastructure maintains a number of central services providing single entry points and guarantees in terms of standards and FAIR guidelines. These services aim to: explore linguistic resources, services and tools available within CLARIN and related communities (through the interface of the Virtual Language Observatory, VLO), find a matching language processing web application for researchers data using the CLARIN Switchboard, to create and publish virtual collections of datasets in the CLARIN Virtual Collections Registry (also connected to the VLO), search for specific patterns across collections of data via the CLARIN search engine (the Search Content).

The big advantage of this integrated system of tools and services integrated in the CLARIN infrastructure is that they can be analyzed and enriched more easily with various linguistic tools (e.g., automated part-of-speech tagging, phonetic alignment or audio/video analysis).

Finally, CLARIN resources might be useful for all SSH disciplines, although the main users are in the fields of linguistics and automatic language processing. CLARIN is particularly aimed at European researchers, but is widely open to other communities (e.g., research outside Europe, citizens, journalists etc.).

OPERAS (https://www.operas-eu.org/)
Standing for Open Scholarly communication in the European Research Area for SSH, OPERAS is the last ERIC to join the ESFRI roadmap in 2021. For now, its main goal is to develop support for scholarly communication in the SHS and to perfect its economic model and community-based governance.

A relevant aspect to report here is that OPERAS is involved in the CCI with libraries and some museums on issues of scientific communication. Another relevant characteristic for our report is the *citizen sciences* component that OPERAS intends to develop. This is partly the case with the European project COESO (https://www.operas-eu.org/projects/coeso/, 2021-2024), which is developing its socio-citizen platform to bring together social actors, researchers and citizens. Within the framework of this project, several pilots involving this type of actor are being developed (e.g., dance and research).

E-RIHS (http://www.e-rihs.eu/)
E-RIHS is a new ERIC in construction (Preparatory Phase in 2017-2020, Transitional and Implementation Phase 2020-2021, transition to Operational Phase will start in 2022). The E-RIHS mission is to deliver integrated access to expertise, data and technologies related to Heritage Science through a standardized approach, and to integrate world-leading European facilities into an organization with a clear identity and a strong cohesive role within the global Heritage Science community. In particular, the focus of the project is on Cultural Heritage conservation and restoration. New instruments, new protocols and new techniques have a decisive impact on heritage science research, enabling improved understanding of heritage objects and sites. E-RIHS ERIC stimulates innovation in large-scale and medium-scale instrumentation, portable technologies and data science.

E-RIHS ERIC promotes good practices and develops or advances methods designed to respond to the specific needs of Cultural Heritage assets, whether material or digital: objects, collections, buildings and sites.
Through interdisciplinary access to the four platforms (E-RIHS ARCHLAB, E-RIHS DIGILAB, E-RIHS FIXLAB, E-RIHS MOLAB), E-RIHS ERIC supports a wide variety of research, from smaller object-focused case studies, to large-scale and longer-term collaborative projects.
The DIGILAB will provide support for Heritage Science data management (following FAIR policies). It aims at providing **virtual access to data** and **tools** for Heritage research; searchable registries of multidimensional images, analytical data and documentation from large academic and research and heritage institutions.
The focus of the Preparatory Phase has addressed legal status and governance/management organization; the provision of effective services is demanded to the operational phase.

1.2.4 Specific European organizations and projects in the area of Cultural Heritage

At different levels, several European programs aim at tooling up Cultural Heritage scientists. Opportunities for funding Cultural Heritage -related projects are spread in many EC programs (see the document 8 *Most Important EU Funding Programmes for Culture 2021-2027*, https://eucalls.net/blog/funding-prorammes-culture). Among the several actions funded by EC or active in the EU domain, we describe here just a small group of projects, whose impact on Cultural Heritage-related subjects is exemplary.

ARIADNEplus (https://ariadne-infrastructure.eu/)
The ARIADNEplus project (EC H2020, January 2019-december 2022) is the extension of the previous ARIADNE Integrating Activity, which successfully **integrated archaeological data** infrastructures in Europe, indexing in its **registry** about 2.000.000 datasets (ARIADNE portal). ARIADNEplus builds on the ARIADNE results, extending and supporting the research community that the previous project created and further developing the relationships with key stakeholders such as the most important European archaeological associations, researchers, heritage professionals, national heritage agencies and so on. The new enlarged partnership of ARIADNEplus covers all of Europe, including leaders in different archaeological domains like palaeoanthropology, bioarchaeology and environmental archaeology and other sectors of archaeological sciences; and touching all periods of human presence from the

appearance of hominids to present times. Transnational Activities together with the planned **training** will further reinforce the presence of ARIADNEplus as a key actor.

The ARIADNEplus **data infrastructure** will be embedded in a cloud that will offer the availability of **Virtual Research Environments** where data-based archaeological research may be carried out. The project will furthermore develop a Linked Data approach to **data discovery**, making available innovative services to users, such as **visualization**, **annotation**, text mining and geo-temporal data management.

3D-ICONS - 3D Digitization of Icons of European Architectural and Archaeological Heritage (http://3dicons-project.eu/)
The focus of this EC project (2012-2015) was to produce digital 3D models of Cultural Heritage artworks/monuments and to contribute those data to Europeana. It aimed at:

- establish a complete pipeline for the production of 3D replicas of artworks, archaeological monuments and historic buildings which covers all technical, legal and organizational aspects;
- create 3D models and a range of other materials (images, texts and videos) of a series of internationally important monuments and buildings;
- contribute content to Europeana using the CARARE aggregation service.

Thus, the project developed and experimented a data processing pipeline covering the entire spectrum from digitization to upload of content on local repositories, with related metadata contributed to Europeana. It covered all aspects of a 3D digitization project from planning and obtaining permissions, selection of methods and tools, data acquisition, post-processing, publication of content online, and metadata capture to making the content available to Europeana.

Many other standard EC projects (RIA, IP) have contributed with programs, research and results concerning Cultural Heritage and museums. This is a well populated set; we cite below just a few of them to mention some specific resources contributed to the community.

V-MUST - Virtual Museum Transnational Network (http://www.v-must.net/)
This EC Network of Excellence project (2011 - 2015) focused on several technologies and methodologies supporting museums' needs. It aimed to provide the heritage sector with the tools and support to develop **Virtual Museums** that are educational, enjoyable, long-lasting and easy to maintain. The V-MUST network organized and hosted a range of activities including conferences, workshops, training sessions, pilots and exhibitions.
Among the resources produced, we mention: an **archiving** service is devoted to digital multimedia content archiving (web interface, enabling semantic search); the **Model Convert** web service for the preparation and conversion of 3D models for online applications, with a specific attention to the needs of online Virtual Museums; the **Render Farm**, a Blender-based rendering service aimed at providing rendering capabilities for computer graphics movie production; the **3DHOP** (3D Heritage Online Presenter) platform for easy publishing 3D content on the Web and interactive visualization of Cultural Heritage data, designed in the framework of V-MUST and further evolved by CNR (Italy) with the support of other EC projects.

4CH (https://www.4ch-project.eu/)
4CH started on the 1st January 2021 for a duration of three years. The project aims to set up the methodological, procedural, and organizational framework of a **Competence Centre**, an infrastructure dedicated to knowledge organization and transfer through means such as training, standardization and inter-disciplinary collaboration. This European Competence Centre will need to adapt to many different conditions such as how Cultural Heritage is managed, the risks that may affect its

conservation and how it may be valorized. Using a holistic and multidisciplinary approach to the Cultural Heritage conservation, it will facilitate coordination between Cultural Heritage Institutions across Europe and provide services and tools to enable preservation and conservation of historical monuments and sites using the latest, most effective technologies with special attention to 3D.

4CH will benefit a range of institutions and other bodies, both public and private, responsible for managing European Cultural Heritage, service providers (Cultural Heritage professionals and SMEs), the creative industries and hospitality sector and heritage agencies, public bodies such as Ministries and decision-makers who inform policies and strategies for conservation, preservation and digitization. The project includes an Advisory Board consisting of high-level experts.
The model of a competence centre like 4CH should be definitely linked to the architecture of a European Collaborative Cloud for Cultural Heritage. Indeed, this represents important distributed expertise resources, crucial to keep thinking and developing standardized processes for Cultural Heritage objects.

Time Machine (https://www.timemachine.eu/)
Time Machine is aiming to join Europe's rich past with up-to-date digital technologies and infrastructures, creating a collective digital information system mapping the European economic, social, cultural, and geographical evolution across times. In the proposed approach, **digitization** is only the first step of a long series of extraction processes, including document segmentation and understanding enhanced by Augmented/Virtual Reality (AR/VR) applications, leading to simulations of hypothetical spatio-temporal 4D reconstructions.
Such **computational models** are key resources for developing new critical reflections on our past and future, enabling new insights for historians, social scientists, creative arts professionals, policy makers, and for the general public, with a significant common denominator: contributing to informed decision-making from everyday life to academic, professional and political matters.

The Time Machine was originally a project finalist aiming to establish a pan-European large project in the Future & Emerging Technologies (FET) program. Due to changes in EU funding the FET Programme was discontinued, but the TimeMachine network organized itself as the Time Machine Organization (TMO). The TMO includes close to 600 organizations from over 40 countries. The TMO network is funded by member fees.

Europeana (https://www.europeana.eu/it)
Europeana is an initiative of the European Union, financed by the European Union's Connecting Europe Facility and European Union Member States, launched in 2009. Europeana's goal is to share and promote European Cultural Heritage to be used and enjoyed by everyone for learning, for work, or just for fun. It provides access to millions of cultural heritage items from institutions across Europe. It does not directly store the content, but provides a web portal to content provided by a network of aggregating partners, which collects the data, checks it thoroughly, and enriches them with information like geo-location, or links it to other material or datasets through associated people, places, or topics.

Among the initiatives described in this subsection, Europeana is the only one aimed primarily at the general audience, rather than to experts only. Thus, it can play an important role in the opening to the wider public context part of the data and knowledge produced and stored on a professional-oriented European Collaborative Cloud for Cultural Heritage.

EIT future new KIC: CCSI (https://eit.europa.eu/our-activities/call-for-eit-communities/2021) Finally, we mention a significant call from the European Institute of Technology (EIT) whose governance is based on various Knowledge and Innovation Communities (KIC), aiming at creating a new one

dedicated to **Culture and Creative Sectors and Industries** (CCSI). This KIC will unite cultural and creative organizations from business, higher education and research centres in a pan-European innovation ecosystem. Thus, "EIT Culture & Creativity will deliver innovative solutions to help the sectors and industries become stronger and more resilient". One of the objectives is also to "bridge regional innovation gaps and harness the power of these sectors to support Europe's cultural leadership".
This call is important because it will identify European consortia and actors to collaborate with and create synergies directly related to the services and tools that will feed the European Collaborative Cloud for Cultural Heritage, in the same spirit we express in section 1.1.

1.2.5 Other related efforts at the international level

Finally, it is worth noticing here also some other important initiatives in the Cultural Heritage field taking place *outside the EU*.

Getty's Research Institute (https://www.getty.edu/research/tools/) and **Conservation Institute** (https://www.getty.edu/conservation/)
The Getty Foundation runs several initiatives, concerning both curation and conservation of Cultural Heritage assets.
In the first domain, the **Getty Research Institute** provides: The Getty Research Portal™, a free online search platform providing worldwide access to an extensive collection of digitized art history texts from a range of institutions (https://www.getty.edu/research/tools/portal/) and a Library Catalogue to access digital collections (https://primo.getty.edu/primo-explore/search?vid=GRI&lang=en_US&search_scope=DIGITAL).
Widely used resources provided the Getty Research Institute are the **vocabularies** including, among others, the Art & Architecture Thesaurus (AAT), Getty Thesaurus of Geographical Names (TGN), and Union List of Artist Names (ULAN) (https://www.getty.edu/research/tools/vocabularies/) . These vocabularies are freely available on the Web using linked open data services.

Conservation-oriented activities are the focus of the **Getty Conservation Institute,** which also provides several digital resources. **ARCHES** is an open-source software platform developed jointly by the Getty Conservation Institute and World Monuments Fund for cultural heritage data management. It is a comprehensive solution for data management, data discovery/visualization, and project/task management. The integrated platform includes: a data management system to manage, define and structure data; discovery and visualization tools to search, report and visualize data (e.g., geospatial data); and project/task management tools (e.g. workflows) to manage sophisticated data editing procedures.
Getty's instruments follow the approach of providing tools to professionals.

Google Arts & Culture (https://artsandculture.google.com/)
It is an online platform of high-resolution images and videos of artworks and Cultural Heritage objects from partner cultural organizations throughout the world.
It utilizes high-resolution image technology that enables the viewer to tour partner organization collections and galleries and explore the artworks' physical and contextual information. The platform includes advanced search capabilities and educational tools.
It is designed by focusing on the large public, to allow anyone in the world to access digital clones of the work of art, with guaranteed quality of the digital representation.

International Image Interoperability Framework - IIIF (https://iiif.io/)

IIIF is an international community aimed at designing open standards for delivering high-quality, attributed digital objects (images and audio/visual resources) online and at scale. It is also developing and implementing the IIIF APIs, backed by a consortium of leading academic and cultural institutions. IIIF offers a way to standardize the delivery of images and audio/visual files from servers to different environments on the Web where they can then be viewed and interacted with (on standard web browsers) in many ways. A set of API defined by IIIF provides a solid foundation to developers allowing the creation of a growing ecosystem of compatible tools, platforms and viewers.

IIIF aggregates a community where to find guidance on best practices, support and feedback for available software and existing resources, a venue to share experiences and projects, and finally discuss future extensions of the API.

IIIF is funded by a 61-member global consortium, and leveraged by aggregators, research institutions, national libraries, archives, museums, software companies, and digital agencies around the world.
IIIF should be an important partner with whom the European Collaborative Cloud for Cultural Heritage should discuss and cooperate. One of the goals of IIIF has been to develop technology which could dissolve the barrier between different museum collections (thus, a common goal with the European Collaborative Cloud for Cultural Heritage), providing tools for cross-collection visualization and inspection.

UK Discovery Projects (https://www.ukri.org/opportunity/discovery-projects-call-towards-a-national-collection-opening-uk-heritage-to-the-world/)
The UK government, under the UK Research and Innovation Council, started the "Discovery Projects" call in 2020, sharing £14.5m of Arts and Humanities Research Council (AHRC) funding to democratize and decolonize the UK's culture and heritage collections.
The five projects selected by the "Discovery Projects" call will form the basis of the 5-year research program *"Towards a National Collection"* aiming to dissolve barriers between thousands of disparate collections (https://www.ukri.org/our-work/browse-our-areas-of-investment-and-support/towards-a-national-collection-opening-uk-heritage-to-the-world/). This is a major investment using digital technology to create a unified national collection of the UK's museums, libraries, galleries and archives to maintain global leadership in digital humanities and arts research (see also at: https://www.nationalcollection.org.uk/).

Among the five selected consortia, *UCL is the coordinator of "The Sloane Lab: Looking back to build future shared collections".* This project will bring Sloane's immense collections, ranging from coins to manuscripts and stuffed animals, which are currently held in a variety of locations and museums, together online for the first time. Therefore, this project is aimed at a full digitization of (dispersed) museum collections.

These initiatives have just started; they are not yet operational in terms of services. But it is worth noticing that the post-Brexit pressure has ignited a strong action on Cultural Heritage themes in UK and EU should be fast in reacting to this action. Indeed, they could, in the long run, be sources of inspiration for the evolution of the European Collaborative Cloud for Cultural Heritage or are interesting to take into account in the framework of future collaborations.

1.2.6 Summary and conclusions

As already introduced at the beginning of this section, the needs of the Cultural Heritage community can be categorized by defining a list of services. We recall here this list and use it to show the features of the initiatives briefly described in this section. The result is Table 1.

Table 1 shows that private companies offer mostly resources fulfilling basic, elementary needs (virtual research environment, storage space for data, processing) or platforms oriented to the large public.
The table shows that no existing initiative is fully addressing the present needs of the Cultural Heritage community. It also shows the need to design a user-driven European Collaborative Cloud for Cultural Heritage to meet the real demands of Cultural Heritage professionals and researchers. Anyway, a number of collaborations are possible and welcomed. The first is with Europeana, which could provide a platform to reach the public. The EC-funded competence centre **4CH** could play an important role in the design of a European Collaborative Cloud for Cultural Heritage, since it could contribute to the identification of competences, requirements, and potential integration of existing services; it could also help in selecting experts who might work on the European Collaborative Cloud for Cultural Heritage assessment of tools and services. Therefore, a collaboration with 4CH would be opportune.
Other initiatives could be strategic for reaching specific communities (ARIADNE for the archaeology domain, E-RISH for conservation and restoration, OPERAS or EIT for innovative and citizen sciences through COESO project or the new KIT CCSI envisioned) and planning integrated efforts.

	HW Data space	HW Computing	Collaboration VRE	Storage & retrieval Catalogue	Storage & retrieval Portals	Storage & retrieval Repository	Storage & retrieval Repository	Storage & retrieval Semantic	Interact w. data Visualizatio	Interact w. data Data	SW support for cloud evolution Software as	SW support for cloud evolution Assessment	Long term access Open	Long term access Data
EOSC	Yes	Yes	Yes	Yes	Yes	Part	No	No	No	Part	Yes	?	Part	Part
OpenAIRE	No	No	No	Yes	Yes	Part	No	No	No	No	Part	Yes	Yes	Yes
EGI	Yes	Yes	Yes	Yes	Part	No	Yes	?	No	No	?	Part	Part	No
D4Science	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes (not all	Part	Yes	Yes	Yes	No
Commercial Clouds: Amazon, Google and Microsoft (Azur)	Yes	Yes	Part	No	No	No	No	No	No	No	Yes	No	No	Part
CLARIN (EC Infra)	Yes	No	Part	Yes	Yes	Yes	Part	Yes	Part (limited)	Part (on text)	Yes	Part	Yes	Yes
DARIAH (EC Infra)	No	No	Yes (limited)	Yes	Part (limited)	No	No	No	No	No	No	No	Yes	Part
E-RIHS (EC Infra)	Yes	Plan	Plan	Yes	Yes	No	No	Plan	Plan	Plan	Plan	No	Plan	Plan
Getty's Vocabularies	No	No	No	Yes	Yes	No	No	Yes	No	No	Yes	Part (voc.	Yes	Part
Getty's ARCHES	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Part	Yes	?	Part	Yes
ARIADNE+ (EC proj.)	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Part	Yes	Part (VRE)	Part	Part
TimeMachine	No	No	No	Yes	Yes	No	No	Part	No	No	No	No	Yes	Yes
4CH	No	?	No	Plan	Plan	?	?	?	?	No	?	Plan	Part	Plan
Europeana	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	Part
V-MUST	No	No	No	No	No	No	No	No	Yes	No	No	Yes	No	No
3D-ICONS	No	No	No	No	Yes	No	No	Yes	No	No	No	No	No	No
ESPADON (Equipex Patrimex, Fr)	Plan*	Plan*	Plan	Plan	Plan	Plan	Plan	Plan	Plan	Plan	No	No	Plan*	Plan*
UK "Discovery Projects"	Plan	No	No	Plan	Plan	Plan	Plan	Plan	Plan	No	No	?	?	Plan
IIIF	No	No	No	No	No	No	Yes	Part	Yes 2D, Plan 3D	Part	No	No	Yes	Part
Google Arts & Culture	No	No	No	No	Yes	Yes	No	?	Yes	No	No	No	No	Yes

Note: **No** = not supported; **Plan** = planned serv; **Part** = partially implemented; **Yes**: fully operational service; **?**: unknown; **Plan***=planned service by cooperating with CNRS TGIR HumaNum

Table 1: Review of Cultural Heritage needs and related support enabled by the major existing initiatives (note: this does not pretend to be an exhaustive and complete list)

1.3 Scientific and technological trends on Cultural Heritage computational modeling and digitization

1.3.1 Cultural Heritage, a complex ecosystem

Driven by the overarching principles of Heritage Science, several initiatives at national and international level in Europe, enabled the development, deployment and exchange of innovative analytical tools for the study of works of art, archaeological objects, monuments and sites, but has also paved the way for a new cross-cutting, multidisciplinary and reflexive approach to comprehensive studies of material cultural heritage, while ensuring its preservation for optimal transmission to future generations.

Cultural heritage institutions are investing in a **comprehensive digital transformation** process during the same period. This transformation process includes both the digitization and online accessibility of Europe's cultural and scientific heritage for educational and preservation purposes.

While the goal of facilitating the use of pan-European resources in a multicultural and multilingual context has largely been achieved, both the digital archiving of scientific data and works and the dematerialization of Cultural Heritage need to be constantly improved in order to integrate technological advances. The pace and scope of technological innovation must be accompanied by corresponding developments in the way we think about Cultural Heritage objects and corresponding new business models.

The digital dimension of Cultural Heritage objects is to date very limited, relevant resources are difficult to pool and still uncoordinated.

Thinking about a **multidimensional, transdisciplinary and distributed tool**, as socio-technical network and infrastructure represent a milestone in the development of Heritage Science.

Cultural Heritage is a complex ecosystem including institutions and actors providing a **continuous production** of data & knowledge on objects from various heritage fields (museums, architectural heritage, archaeological heritage, paleontology, etc.) and materials of different nature (paintings, metal, wood, stone, concrete, glass, organic materials, etc;) and texts and objects of intangible Cultural Heritage.

The study, the conservation and the dissemination of Heritage Objects requires a panoply of **actors** playing in a very large spectrum of **activities**. Three main areas of expertise must be mentioned:

- **Heritage Objects** - or Boundary Objects - (artworks and other heritage objects, monuments and archaeological sites). This area has a strong **SSH** footing and includes historians, archivists, art historians, archaeologists, sociologists of science, museologists, heritage anthropologists, and curators of cultural and natural tangible heritage). This area provides the main base knowledge about the historical and socio-cultural value of heritage objects, taken into account by the applications (conservation, restoration, dissemination) driven by cultural heritage institutions (museums, archaeological sites, etc.).
- **Materials** - (specialists - coming from material and environmental sciences: physics, chemistry, geology, etc.) - in physicochemical analysis and in multi-scale 2D and 3D imaging, with combined know-how in both using and developing instrumentation and in processing data on the heritage objects and their evolution. The joint technological advancement in analysis and representation techniques, including the use of imaging tools and real-time observation.
- **Data** (ITC, computer science, engineering, data science, etc.) focusing on defining methods for data production, processing, sharing, interconnection and analysis. This area is today invested in

multidimensional digitization, geometric and visual reconstruction, modeling and simulations for structural and acoustic studies, development of information systems for historic studies and conservation-restoration, virtual environments for dissemination, semantic-driven analysis, classification, retrieval and interlinking of digital resources on the web, etc.

This is the typical landscape in which, towards a European vision, the challenge of the digital transformation in Cultural Heritage must be addressed. This implies a cross-cutting approach, which ranges from a strong instrumentation effort to the methodological renewal of the scientific and professional practices by a strong integration of Digital, SSH together with Experimental and Environmental Sciences.

1.3.2 Heritage Sciences in the Digital Transformation

In recent years, research laboratories, Cultural Heritage institutions and IT companies have actively contributed to the development of the interdisciplinary field of digital cultural heritage in Europe. First by organizing scientific events on the subject, then by participating in several national and international research projects dealing with scientific, methodological and technological aspects of the acquisition, modeling, sharing and dissemination of digital data for Cultural Heritage knowledge, preservation and dissemination. These initiatives, in conjunction with many other collaborative projects and events, have enabled the international community, particularly at the European level, to address together essential issues related to the (1) formalization of knowledge in the fields mobilized by Cultural Heritage research, (2) traceability and interoperability of data processing, on the (3) implementation of solutions for networking, sharing and exploring heterogeneous datasets, and on the (4) design, development and testing of methods and tools for developing 2D, 3D and 4D representations of current and (hypothetical) past states of objects, monuments and sites.
This digital transformation has many implications for heritage science, as it opens up new sources of data while stimulating a renewal of research questions and methods. Although the entire community now recognizes the importance and potential of digital data, it's important to take a step back to highlight the methodological limitations associated with the mechanisms of its production in the specific context of heritage research, especially with the aim of linking it more closely to scientific knowledge production. To do this, we must first trace two major changes that the digital transformation has brought about in the last twenty years. The first ranges from the digitization of texts to the processing of multidimensional digital data; the second from individual research on an object of study to the development of web platforms that promote collaborative and participatory approaches. At the intersection of these two paths, each researcher's daily practice of collecting, categorizing and storing data is increasingly linked to the construction of digital environments that multiply the contexts of analysis and interpretation by introducing new methods of co-producing knowledge.

During the opening ceremony of the first G20 Summit of Ministers of Culture this year, it was pointed out that *"it's time to introduce a digital humanism in which archaeologists, anthropologists, architects, historians, philosophers, lawyers, neuroscientists and psychologists work side by side with chemists, physicists and computer scientists to define a new semantics to understand the complexity of reality"*[1].
Heritage sciences make the confrontation between material objects and multidisciplinary studies a field for the production of knowledge. This is a privileged framework to analyse the methods of investigation and collective interpretation of facts, objects and phenomena that lead to the production of new scientific and cultural resources, but also to support the transition between an (almost unconscious) production of raw data and a conscious production of semantically rich and

[1] This sentence was spoken by the director of the Egyptian Museum in Turin, Christian Greco, during the opening ceremony of the first G20 Culture Ministers' Meeting in the Colosseum Arena on 29 July 2021

collectively produced **digital commons**, our tomorrow's heritage. It is about establishing a conceptual and informational link between physical objects and multidisciplinary gazes in order to define a new semantics that makes it possible to understand the complexity of reality by linking the material dimension of heritage objects to the mechanisms of scientific knowledge production: shifting the cursor of digitization from material objects to the knowledge mobilized to understand them.

1.3.3 The vision underlying the design of a European Collaborative Cloud for Cultural Heritage

The vision underlying the design of a European Collaborative Cloud for Cultural Heritage should be implemented at the cross–section of three major scientific and technological trends (see next section) that shape the state of the art in computational modeling and digitization in cultural spheres:

- **Towards digital twins** of *heritage objects.* Moving from fragmented and episodic digitisations to "permanent" digital counterparts of heritage objects, able to progressively integrate new data acquisitions, analysis, and enrichment relating to the multiple facets of the study, conservation and dissemination applications;
- **Towards a digital continuum** of activities. Moving from the amount of non-memorized daily activities on Cultural Heritage objects (and related data) produced by curation or study activities of several actors, to a digital continuum, which is able to ensure their traceability and interoperability with the heritage object's digital twin;
- **Towards a digital ecosystem** interconnecting actors, activities and heritage objects. Moving from fragmented and individual data production and curation, to a sophisticated socio-technical system progressively built as a distributed, adaptable and open platform. This platform should offer a number of tools, designed in compliance with specific requirements of the Cultural Heritage communities, enabling an active cooperation of Cultural Heritage professionals over the digital twins. Encoding and archiving knowledge should be supported with easy-to-use interfaces, capable of linking digital counterparts of Cultural Heritage objects to the continuum of activities for their curation, within the framework of the construction of a European social network connecting scholars, curators and the public to common objects of knowledge and interest.

The introduction of a paradigmatic shift in the acquisition and management of Cultural Heritage data (digital, multidimensional, based on multiple and interconnected media, collectively semantically enriched, networked and open) will be a direct result of the implementation of this vision. It must be accompanied by stimulating an international dynamic for producing some representative **digital commons** on European Cultural Heritage objects, which should ignite the population process of a new European Collaborative Cloud for Cultural Heritage and build some reference examples of good practices. This cross-cutting objective, to be articulated with strategic partnerships with on-going initiatives at European level, should stimulate the emergence of new European collaborative study scenarios, to which a few examples are listed below:

- **Gothic architecture** of cathedrals, monasteries and churches in Europe, which has distinctive features in certain regions due to local influences (itinerant craftsmen transporting their ideas between cities, regions and countries) and whose origin and evolution of basic forms (e.g., arches) result from several stages of development that do not progress at the same rate or in the same way in all countries.
- Heritage of **Greek and Roman civilizations**, whose remains can be found in many EU States and are the focus of many museums (with common conservation and fruition needs).
- **Industrial heritage**, characterizing 19^{th} and 20^{th} century history of several EU nations, which contributed to the European industrial revolutions and urbanization, which provides considerable problems of conservation, recovery of lost heritage, and related storytelling.

The introduction of a European framework (infrastructure, actor networks, platforms, tools, etc.) for the creation of genuine digital commons could play a fundamental role not only in addressing the challenges posed by the analysis and massive correlation of data relating to material objects that are spatially distant but close in their characteristics (typologies, styles, rules of composition, states of preservation, etc.), but also and above all in examining the complex relationships that researchers associate with their knowledge objects. The point is to create a territory of digital data that can be explored by following the paths of physical and semantic similarities that link objects of study, scientific questions, study protocols, instruments, descriptive terms and individuals to form new research communities that transcend geographical, institutional and disciplinary boundaries.

1.3.4 Scientific and technological trends and gaps

To position this vision in the scientific and technological literature, several facets of a fragmented progress in the digitization of cultural heritage need to be analysed in an integrated way. The digital turn in recent decades has had many impacts on related research activities. It provided the actors with new data types, ranging from visual and non-visual data and metadata (see section 2.1 on data types), leading to a renewal of research and development questions and methods. Following the emergence of digital libraries solutions for managing documentary resources, the range of tools for generating digital data on cultural heritage objects has been gradually enriched by new digitization methods [3] and processing and analysis algorithms [4] . In the fields of art history, architecture and archaeology, 3D imaging has been used to reconstruct previous or hypothetical states [5], [6], in particular to perform simulations [7] [8] [9] or to perform anastyloses that would have been very complex without the use of virtual environments [10]. For conservation and restoration activities, digital approaches are also extremely valuable, whether for conducting condition reports or for monitoring a particular object over time [11] [12]. Apart from the trend towards producing and having better digital twins, the Cultural Heritage community needs a really strongly integrated complex including the design of novel integrated activities in the digital continuum, together with a trend towards a digital ecosystem including all concerned actors and stakeholders.

Towards Digital twins. After decades of work on digitization of text, images, videos and sounds, in recent years, 3D imaging and, in particular, reality-based 3D reconstruction have become indispensable for the documentation of cultural heritage. 3D imaging has the triple capacity of capturing the objects under study, improving the analysis of their morphological complexity, and reproducing the result of their interpretations. In this context, approaches based on photogrammetry and computer vision [13] have revolutionized the creation of 3D digital representations of cultural heritage objects through a democratization process that ranges from the use of low-cost image sensors [Kirchhoefer et al., 2011] to their large-scale deployment in crowdsourcing scenarios [15]. These methods are gradually becoming part of the daily monitoring of cultural heritage objects in archaeological, conservation, and restoration studies. Research on the creation and processing of digital 3D representations has been certainly productive, proposing new and advanced algorithms on geometry and appearance acquisition, on the measure of their fidelity, on handling gigantic digital models with high data complexity, on data compression and data simplification, on multi-resolution encoding and progressive transmission, etc. Research has also advanced in the field of interaction with the digital Cultural Heritage content (see section 2.3 "Interaction with Cultural Heritage data/content"). But moving towards the digital twin entails the need for efficient digitization, data archival and semantic enrichment processes.

Concerning Digitization, computing precise local fidelity bounds in huge reconstructed 3D models is still an open problem with intriguing research challenges (see section 2.1.3 on Data quality and validation methodologies). Another avenue of future work is on improving the appearance attributes of the digitized artworks, achieving high quality materials mimicking the real objects and being able to measure the fidelity of the digital materials (see section 2.1.5 on further research on digitization). Digitization of large environments is frequently limited by the lack of physical accessibility to certain areas of the monuments [34].

On data archival, a recent review [32] analyses 3D viewers and repositories like 3DHOP and the Universal Viewer, Smithsonian 3D, Three D Scans, CyArk, Europeana, EPOCH, CARARE, NASA 3D, Sketchfab, MyMiniFactory, Blendswap, 3D Warehouse, TurboSquid, ShareCG, 3DExport, Free3D, and others. These repositories, such as ShareCG, TurboSquid, CG Trader and Yeggi, typically lack data provenance and metadata. The authors also note that existing repositories are not supporting a range of useful features like measurement tools, the ability to link to archival records, methods to track online and offline use, or the visual analysis of the 3D model over time (supporting changes in material, space and use, also having a timeline-related model information could be highly useful to experts and users).
Using Cultural Heritage models and data for long-term archive is another uncharted field (See section 2.1.4 on Data preservation). Long term archive of monuments and artworks will require obviously long-term access to digital Cultural Heritage data, but also a kind of continuum (like a 4D digital whole) constituted by space-time sequences of connected and related digital twins with a common narrative: a sequence S of a number of different digital twins physically located in different places in Europe but linked together by a common history (art style, sculptor, etc.), but in some cases also being acquired and reconstructed every some years to observe its evolution, erosion, degradation etc. Creating, maintaining, enriching and preserving such a 4D digital whole could certainly be one of the challenges of the near future, to *move from fragmented and multiple digitization to "permanent" digital counterparts of heritage objects, able to progressively integrate new data acquisitions, analysis, and enrichment relating to the multiple facets of the study, conservation and dissemination applications.*

Towards a Digital continuum. In order to memorize and interconnect digital resources coming from daily activities carried out by several actors on heritage objects, a first problem concerns the traceability and interoperability of digital data creation processes [17], in a context where the data generated are often guided by constraints of instrumentalization and cognitive processes that we cannot adequately trace, and where their use for scientific purposes requires a detailed understanding of their origin and nature. Several works have addressed these issues [16], [18] and discuss the definition of metadata and paradata that must also be considered at each stage of data creation and/or transformation [Dudek, Blaise, 2017]. The question of formalizing knowledge models is related to, but different from, the question of traceability, as it can only be solved by observing and analyzing the practices of the different disciplines that come into play. Various approaches from the field of knowledge engineering have provided methods [20], but the challenge here's to formalise the various knowledge domains associated with the multidisciplinary nature of a broader and complex domain (see section 2.2 on coupling and integrating digital representations with metadata and semantics). This topic can benefit greatly from work in information science, particularly in formal ontologies. Ontologies and knowledge representations are commonly used to structure terms and concepts related to specific knowledge domains. The conceptual model CIDOC CRM [21] is specifically tailored to the cultural heritage domain. Of particular interest for navigation and inference are approaches from the field of knowledge modeling, such as those proposed by [22] or [23], which are based on graph-oriented databases and enable rapid browsing of a set of data and their relationships. By following this perspective, even though knowledge engineering approaches introduced solutions for aggregating, implementing and investigating the evolution of data, these approaches remain difficult to apply in an effective way [24], especially in the specific context of cultural heritage, where workflows often involve subjective human decisions, poorly formalized knowledge, non-explicit research protocols, highly individualized skills.

On another perspective, if today we are witnessing a massive but scattered production of digital resources of cultural heritage objects, the growing mass of uninterpreted data reveals an essential need for innovative methods to move towards a centralized mass production of semantically enriched digital resources within multidisciplinary research perspectives. The recent transition to mass production of spatially representable images and models in 3D, which also enables 4D acquisition and visualization of real objects [25], has recently raised new questions at the interface between the management of data masses and their semantic enrichment, and their automatic classification for research purposes by content similarity [4]. Semantic enrichment of reality-based 3D representations ranges from methods for linking semantic tags to 3D models [26] to approaches for structuring heterogeneous data around a building information model [27]. Joint analysis of spatial and temporal data includes the study of object transformations or the modeling and representation of categorized events [28]. Finally, the measurement of similarity between different digital objects has been an active research area for several years, from the introduction of image analysis methods to computer vision and recent approaches to the analysis of 3D models. These methods provide interesting results in indexing general multimedia objects [29], which are also applied in the field of cultural heritage [30] [31].

The use of Machine Learning (ML) and Deep Learning (DL) techniques in Cultural Heritage is still limited [33], although it will certainly grow in the near future, with applications ranging from chronological classification of ancient paintings to the automatic annotation of visual contents in ancient manuscripts or artwork analysis. All in all, ML and DL will have to be used with expert post-supervision, to detect and avoid their unavoidable errors and misclassifications. These approaches should also be used for aggregating multi thematic observations around the semantically structured digital representations, by introducing autonomous mechanisms of connections between heterogeneous resources to *move from the amount of non-memorized daily activities on Cultural Heritage objects (and related data) produced by curation or study activities of several actors, to a digital continuum, which is able to ensure their traceability and interoperability with the heritage object's digital twin.*

Towards a Digital ecosystem. In a broader context, collaborative science [Boyer-Kassem et al., 2018] has gradually invested in a large number of fields by introducing new links between data production and scientific analysis. With the web, international open science has experienced unprecedented development, paving the way for new subfields (data science, artificial intelligence, etc.) and roadmaps for the world's major research organizations. However, there is still a significant imbalance between the sharing of a mass of raw data and the collaborative production of semantically enriched data that needs to be questioned today. Indeed, in recognizing the value and potential of digital data, it is necessary to take a step back to recognize the methodological limitations associated with the mechanisms of its production within the scientific and professional framework.

From a pure technical point of view, isolated solutions on digital analysis, annotation, browsing of different media (from text to 3D imaging) should be merged into systemic integration (see section 2.3 on Interaction with Cultural Heritage data/content) and experimentation frameworks towards socio-technical environments for collaborative work (see section 2.4 on potential uses for museums or Cultural Heritage institutions). As a fact, we now have a universe of partial repositories and tool systems, many of them aimed at museum visitors and restoration and art experts, some of them also focused on the research community, and most of them being isolated and partial initiatives. However, the Cultural Heritage community lacks a unified framework for long-term access to digital Cultural Heritage data, and a set of tools shaping a single interconnected system, an ecological complex being really useful to stakeholders and usable by them. Data obsolescence and date reuse remain unsolved, data provenance is missing, data coherence and integrity is usually not considered. In any case, the analysis of which new features could and should be included must come from the user community requirements, including galleries, libraries, archives, museums, heritage scholars and heritage communities.

Data preservation is not just ensuring long lasting physical retention of the digital data (using backups, replicated storage, use of permanent data memories, etc.). Data preservation should also avoid data losses due to the obsolescence of data formats and data access applications (see section 2.1). This is a real technological challenge, due to the fast evolution of data types and storage systems. Technological development should not only be devoted to the design of innovative systems to manage novel digital models and data, but to guarantee the survival of old data by keeping them alive and permanently accessible.
Finally, as already observed, long-term access to public infrastructures on digital Cultural Heritage data is not presently supported, and this should be an important initiative in the next few years. This includes long-term mechanisms to avoid data obsolescence and to ensure date reuse. This is not trivial, and will probably be one of the essential scientific and technological areas of work during the next few years, to progressively build *a distributed, adaptable and open digital ecosystem including tools and services designed in compliance with specific requirements of the Cultural Heritage communities, with easy-to-use interfaces, capable of linking digital counterparts of Cultural Heritage objects to the continuum of activities for their curation, within the framework of the construction of a European social network connecting scholars, curators and the public to common objects of knowledge and interest*.

1.4 Assessment and sustainability

The assessment of any human endeavour is a critical and complex activity. Research is a domain where validation and assessment are also key actions. We have a consolidated experience on evaluating research results (e.g., peer reviewing of scientific papers or by means of numeric indicators), and a lot of current debate on possible changes of the modalities used to perform this assessment.
Research projects follows the following track: in any funded (EC) project, the consortium is asked to produce a solid evaluation of the results produced. Any action directed to the design of a European Collaborative Cloud for Cultural Heritage should follow good assessment practices, or even drastically improve those practices. We will consider that the assessment would be strictly related to sustainability.

First, we should clarify what we include in the **results of a project**.
Scientific papers are among the main results of a project. But this outcome follows specific and consolidated assessment policies, which are still under debate.

We are more interested here in discussing other categories of potential output: **Data, Software Tools, Online Services**. The policies for validating the quality of these categories of results are less consolidated than the ones used to assess quality of publications, which raise the following questions:

Who? - Who is entitled or even partially responsible for the evaluation. In the case of standard research projects funded by EC, in most cases the evaluation is supervised by (or directly commissioned to) partner(s) of the consortium. This might create a conflict of interest, especially if the test and validation is performed with a limited contribution of users external to the project consortium.

When? Assessment usually occurs in the very final phase of a project. The **design-implement-test** cycle is usually performed in a short period of time (usual projects' life span is 24-48 months) and the testing phase is usually performed in the very final period, with a duration of 4-8 months. The main issue is that the assessment is done in the very early life of a SW tool, usually as soon as the tool is released. Thus, it is more a test of operation of the tool than a measure on how well it fits the community expectations. In real life, a tool becomes public at the end of a project and demonstrates its utility and impact many months or years after the end of the project. Thus, the single project lifetime (M1, M36 or M48) is not adequate for assessing the effective impact of a tool. It should be possible to apply an **all-life assessment** on data or tools, taking into account their performances during the entire life cycle of the asset.

The situation is even more complicated with longer term actions, such as the design and implementation of an infrastructure, which should be running for many years. In this case we may have a longer duration of the supporting funding action (thus enabling an all-life assessment), but the potential conflict of interest is still holding (the project consortium should not act as self-evaluator).

The limited time span of most projects is the reason for another problem: after the completion of a project, results produced (even when successful and used by a supporting community) risk to be abandoned by their creators. Curing a dataset, maintaining and extending a SW tool or a web service is an activity requiring human effort, and thus needing dedicated funds. It is unfortunately common that project's results are discontinued immediately after the project's termination, causing a huge loss for both the funding institution and the creator institution/consortium. This loss has to be prevented and is strictly related with the assessment theme.

Supporting maintenance of digital resources is a complex task (evolving data types, obsolescence of architectures/devices/libraries/languages). Funding institutions cannot ensure to dedicate funds to all the results produced by the financed calls/projects. This is both because research funds are limited, and because the decision on what has to survive should be based on an **evolutionary approach** (only the best should survive, after proving to be strongly beneficial to a community of users). Therefore, to prevent the loss of project results, to keep alive the success stories and, possibly, to enforce them we should be able to run serious and selective assessment tests. This will also have an impact on **sustainability**, because it will become known what is more important to preserve, i.e., the more focused, the easier to preserve.

Therefore, it is crucial to establish a common policy and related instruments to understand: (1) which are the project results really backed by an **engaged community of users** and (2) the related **impact** over the domains of application. While the first criteria could be seen primarily as an evaluation based on numeric indicators, the second would be more lying towards subjective indicators. The overall assessment work should be a mixture of both.

To solve this issue, a European Collaborative Cloud for Cultural Heritage should go beyond the functionalities needed for basic data archival and services/tools provision. It should also support:

- Instruments enabling to produce information on the **level of access** and **reuse** of each data asset stored in the cloud (access, visualization, reuse in other researches or projects);
- Instruments enabling to gather numeric information on the **level of vitality** of each service/tool installed on the cloud, including both numeric data characterizing complexity and frequency of updates operated over the tool (e.g., the ones popularized by GitHub or similar platforms: number of source lines, number of contributors, commit frequency, etc.) and numeric indicators on the community of users (number. of downloads, number of scientific papers citing the use of the tool, other automatic users' satisfaction indicators, etc.)
- Instruments enabling more **subjective evaluations** of services/tools/assessment of user satisfaction via questionnaire or interviews), resembling the peer-to-peer evaluation that we are used to for publications. Independent experts should play an important role in this activity.

Once we have a solid assessment policy, we could also produce answers to the main initial question: who will pay for extending the life time of project results, i.e., the **sustainability** of the project.

In other application domains, this is the role of private companies. They should have excellent knowledge of the market and a clear view on the leading technologies which will lead to successful technology transfer from the academic domain to the final customers.
This is unfortunately not the case of the Cultural Heritage domain that has a largely insufficient commercial appeal for most industrial companies. The experience of the last twenty years has demonstrated that the Cultural Heritage domain is not able to endorse technologies and steer their evolution to the specific needs of this community without a public investment.
Therefore, it is not just a matter of funding the creation of a European Collaborative Cloud for Cultural Heritage (and related services), possibly with multiple projects, but also of its maintenance on time. Funding institution(s) should provide instruments to finance the maintenance and evolution of the successful components (the tools/services performing better according to the evolutionary model). Therefore, the Cloud should have funds available (on a medium-term scale, e.g., 5-10 years, renewable) to ensure both to cover the cost of the Cloud (managed by a consortium) and the cost of maintenance of successful tools/services (contributed to the by external partners, either academic or industrial, in the framework of other EC projects).

Perhaps we should consider working together with multiple Cultural Heritage institutions in the development of the project, and maintain these partners to manage the platform after its launch. For this to be at all feasible a management and data budget covering at least several years is necessary (and should be included in our estimates) AND the platform should be open source, so that a community of participants can grow. We need a separate plan on how to build this community on the basis of early adopter partnering Cultural Heritage institutions.

Once the European Collaborative Cloud for Cultural Heritage will be consolidated (populated by a large number of digital assets, containing many useful tools and services) sustainability could also be helped by **pay-per-use policies**. But, knowing the state of Cultural Heritage institutions in the EU, these pay-per-use policies should be mostly directed to commercial/private entities (e.g., creative industries, publishing industries active on education books, etc.) rather than the public Cultural Heritage institutions. We should also consider that the success of the Cloud will depend on the contributions of many Cultural Heritage institutions, since many of them will contribute substantially in-kind with data. EC should probably not implement a business model, in which these institutions are required to pay for service.

Finally, we underline that preservation and maintenance of digital resources is not just a technical problem, but also a governance issue, which will require work on both economic and managerial sustainability. Some of these themes will be further discussed in section 3.4 *Governance, Business Model and Sustainability*.

Moreover, a strong emphasis and support on all-life-long **dissemination** and **training** could also affect **sustainability**. This theme is also discussed in section 3.4 *Governance, Business Model and Sustainability*.

We should also mention the importance of **environmental sustainability.** Developing a green platform, or a platform that will not increase substantially our carbon footprint is another topic that is discussed in section 3.5 *Contributions to Green Deal*.

PART 2 - CURRENT STATUS OF THE CULTURAL HERITAGE DOMAIN - DATA, REQUIREMENTS AND VISION

2.1 Data types for Cultural Heritage-related applications

An important characteristic of most Cultural Heritage activities is the use of many different types of data. Conversely to many other application domains, Cultural Heritage does not define a single preferred data type, any of them have specific uses and are considered optimal for some specific applications.
Therefore, an infrastructure aimed at supporting the wide spectrum of Cultural Heritage applications and data uses should fully support any of the data types used.
Data has many different meanings and incarnations in Cultural Heritage. We describe here the characteristics of the different types by endorsing a first distinction: **visual** and **non-visual** data.

Representing the physical essence of Cultural Heritage assets (artworks, buildings, tools, etc.) is a specific need in Cultural Heritage. We need digital representations able to encode representation(s) of the shape and the appearance of our assets. We call those types of representation **visual data**, to underline the immediate and intrinsic visual counterpart of the data content.

But representing Cultural Heritage assets is much more than encoding just the physical essence. Other information has to be collected, represented and linked to the asset. We call these other data **non-visual**, and we do not mean just text. Data which allow us to characterize the materials are an example.

Moreover, data can be created to support study and research (e.g., the digitization of an asset, produced as a preliminary phase of a study or of a conservation study), but can also be the product of the research (data produced by specific research, i.e., results of scientific investigations; or data which encode the process adopted to reach a specific insight, which we need to store and archive, to make research reproducible).

Finally, data comes at different levels of quality and fidelity. Since those data are important instruments in Cultural Heritage research, it is important that the quality of the single data could be assessed and reported, to allow scholars or professionals to understand the level of trust of each specific digital asset. In addition, the source of all data should remain visible, including when the data is a part of a larger body consisting of data from various sources. This will enable combined research efforts from different sources, completing knowledge on artists, locations, techniques etc. by working together as the European Cultural Heritage sector, and it will have many other advantages. Data sources could be protected on the blockchain, so that even when content is dissolved in other content its creator would remain visible.
Data comes from different places. Museums, private collections, dealers. They come from institutions involved in research projects, and from those outside the project.

2.1.1 Visual Data

Many types of visual data are used extensively in Cultural Heritage. We list here the main subclasses and present a very synthetic description, together with some bibliographic references.

Standard 2D Images
Images are the more common visual media. They have been part of art history and archaeology from the very beginning, initially employing the analogic, printed version and, more recently, digital supports (either digitally native images or scanned from old prints/slides).

While images are fully integrated with the web and HTML since the *World Wide Web* birth, few aspects lack a standard solution for archival and visualization purposes. Most of the images produced nowadays are very high-resolution. High-resolution images are now a commodity resource, given the impressive evolution of digital photography (just to mention a single example, recent off-the-shelf smartphones provide 20-80 Mpixel cameras). Moreover, the availability of tools that allow aligning and stitching image patchworks supports users in reaching huge image resolutions.
Visualization on the web of high- or huge-resolution images can be tricky but doable, using specific methodologies (data compression, efficient progressive data transmission). Examples are the multi-resolution approaches based on tiling and hierarchical image representations [35], [36]. Another important and critical issue could be protecting the data (image watermark technologies [37]).

Digitized maps, manuscripts or historical documents are often digitized and encoded as a specific type of content for 2D images. Many historical documents come in the form of hand-made documents (historical maps, drawings, letters, reports, etc.). In some cases, these documents are digitized and treated as 2D images (we can also apply digital restoration techniques to enhance readability and reduce the effects of degradation). Conversely, documents containing mostly written text are often converted in text format (using handwritten characters-to-text technologies) and thus treated as non-visual data.

Multi-spectral images
These specific 2D images depict the light reflected by a surface while sampling only a specific wave-length [38]. These are quite common in Cultural Heritage, especially for investigation and conservation purposes. Some examples are infrared images (allowing the detection of under-drawings) or ultra-violet images (which disclose the presence of patinas or layers of biological substances, and faded or modern paints).
If we consider just their digital information content they can be treated as standard images, usually encoded using a single channel (grey levels). But they require specific management because we usually need to manage them at the analysis/visualization stage as a group of images and using specific visual analysis procedures.

Reflection Transformation Images (RTI)
Relightable images (usually named *Reflection Transformation Images - RTI*) are becoming an increasingly used technology to acquire detailed and interactive documentation on quasi-planar objects characterized by complex light reflection attributes [39]. The advantage of this representation is the possibility to interact with the image: we can change the direction of the light incident over the object in real-time (i.e., at visualization time) to inspect fine details of the objects' surface. The visual quality and fidelity supported by this media are impressive, in many cases, superior to what we can simulate with 3D models.
RTI images' acquisition is quite simple, requiring a calibrated lighting system and shooting multiple photos from a stationary camera position under variable lighting. Those input images are then processed to produce a single RTI image. The images' lighting information is mathematically synthesized to generate a mathematical surface reflection model for each specific parcel of the surface reflection, enabling users to re-light the RTI image interactively and in real-time. Thus, RTI images encode, for each pixel image, nor the RGB value but a *function* able to return the surface's colour given a specific direction of the incoming light incidence.

Panoramic Images
Panoramic or 360 degrees images (or panoramic videos) add a different interaction opportunity, allowing users to navigate and interact with these visual assets [40]. Acquisition ease and speed, together with the richness of details granted by the high-resolution support, makes panoramic images an ideal medium for Cultural Heritage, very well-fitting fruition and dissemination needs.

The acquisition of panoramic videos is also straightforward, either from multiple poses taken by a camera positioned over a tripod or using specific devices (based on multiple video cameras).

3D representations

3D representations also become quite common in Cultural Heritage [41]. They usually represent the visible external surface of real 3D objects. Two classes of models are used:

Sampled models, usually produced with active 3D scanning (laser-based systems or systems using structured light) or with photogrammetry approaches (production of 3D models from set of 2D images);

Modelled representations, produced with user-driven modeling systems designed for 3D CAD modeling and computer animation applications (e.g., Blender, Maya, etc.).

Sampled models give much more control on the accuracy of the representation than *modelled* representations. Conversely, the latter are more common for applications designed for the public (e.g., to produce videos or virtual reconstructions).

CH applications often require to encode not just the *shape* but also the *appearance* of the artwork (the surface colour or the more sophisticated surface reflection properties). The shape is represented either using a triangle-based encoding, or using point clouds. Appearance is often encoded by adding textures to the shape representation.

There is a pressing need for platforms supporting easy and free publication on the web of 3D models. SketchFab [42] is a recent commercial solution, supporting automatic web publishing and a nice and easy to use interface.

Animated/deformable 3D models are used in Cultural Heritage to add an interactive behaviour to the artefact. They are complex models which allow encoding both the shape and the functioning of an artefact. Therefore, not just how it looks but also how one could operate/manipulate/act with it.

Depth images (or *depth maps*) are another type of visual data with a predominant 3D interpretation. These are usually the raw result of a 3D sampling performed with an active scanning device: an image where each pixel encodes a point sampled in the 3D space. We can render a depth map as an image (thus, rendering in false colour the distance from the observer), or we can convert it into a cloud of points. These data are usually raw data, produced as intermediate results of a 3D digitization process.

Terrain Models

Terrain models are commonly termed as 2.5-dimensional data. They are quite common in geographic or land representations and are often used to represent the context of Cultural Heritage discoveries visually. These data are managed with GIS approaches (when we have to characterize the use or meaning of specific parcels of land) or as standard 3D data.

CT data

Data produced by tomographic machines (either of industrial or medical derivation) are often used as an instrument to see behind the external surface. Very common applications in Cultural Heritage are to unwrap virtually mummies, or to discover the content of sealed vases, or to detect damages and fractures in the interior of solid materials. Those data come with specific formats (pile of 2D images); in some cases, their content is segmented and encoded with standard 3D representations.

Videos

Videos are a standard media also in Cultural Heritage, probably more used in dissemination than in conservation or study activities. Video technology is extremely consolidated. Video footage can be produced by grabbing a real scene with cameras (now on any smartphone) or by producing synthetic videos using computer animation tools. Widely used platforms exist for web-based video upload and streamed access.

Beyond Visual – Sound

Sound is an important component for the simulation or representation of Cultural Heritage spaces. A silent 3D scene is not realistic nor sufficiently immersive in a virtual reality context. Thus, the sound should be taken into account while producing sophisticated visual products, as it is the case of computer games which pair visual and sound contents.

Data produced by diagnostic devices

Data is also produced by many diagnostic instruments, used routinely for assessing the conservation status of an artwork or for producing evidence and insight on its characteristics. Those data are very often 2D images, where the pixel intensity represents the value of the sampled attribute. In other cases, we have tabulated results (usually presented with graphs). Thus, the output of these investigation devices falls over standard data categories; the specificity here is that we have important meta- and para-data which should be paired to the visual or tabulated data (see next subsection on non-visual data).

Common issues with visual data

There are some issues which are common to most of the visual data types we have briefly presented. The first is **data complexity**. The resolution supported by modern 2D and 3D acquisition devices makes the resulting dataset very heavy, with issues in transmission and interactive visualization. But more than 20 years of research on data compression, data simplification and multi-resolution encoding, progressive transmission gives the instruments for implementing tools able to manage even extremely dense representations in real time and on commodity devices. But it is mandatory to use those technologies in the design and implementation of efficient data access solutions.

The second is **interaction management** and **ease of use**. To facilitate the access to multiple types of data it is mandatory to provide consistent user interfaces and manipulation metaphors. The main reason for not using a digital instrument is often linked to a not proper design of its user interface, rather than to the effective fit of its functionalities. This is extremely critical for not highly technical communities. The interface of a 3D visualizer should be consistent with the one of a 2D image inspection tool, or better they should be designed in a consistent manner.

Data **obsolescence** is another issue, since visual technologies evolve very fast (see subsection 2.1.4). Here an important role is played by standardization efforts. The multimedia context is characterized by a very fast technological evolution. There are many different **data types** proposed for each media, some of them are endorsed by a standardization committee, others are consolidated as market-driven formats. It is not easy for Cultural Heritage users to drive decisions in such a complex technological domain. The instruments provided by the European Collaborative Cloud for Cultural Heritage should hide as much as possible this complexity and should endorse all leading standards. Among the various activities in this domain, the work of the **International Image Interoperability Framework** (**IIIF**, https://iiif.io/) is strongly relevant and has a clear impact in the Cultural Heritage context. IIIF produces directives and components, including specific viewers (more consolidated for 2D images, in course of definition for 3D data). IIIF resources and experience should be taken into account while designing the visualization mid-level components of the European Collaborative Cloud for Cultural Heritage.

Finally, visual data is often inspected in an isolated manner. But the insight process often needs to inspect and analyse many media, thus a trend for the future is to design and support visual navigation and analysis approaches which ***use multiple media in a single common context***. There are already some approaches presented in literature and is a promising subject of future research [43].

2.1.2 Non-visual data and metadata

Linked Data

As more and more Cultural Heritage data is published on the World Wide Web (WWW), the data models supporting this have become more and more popular. The model there, the "Layer Cake Model" [48], contains several levels of abstraction and standards for document and knowledge representation [49]. On the lowest level, there is the Unicode character system for representing text in different languages and writing systems, and the URI/IRI identifier system for minting globally unique identifiers for web "resources", i.e., anything that one can represent knowledge about. Next there is the XML layer for representing documents and data, the "lingua franca" of the Web. The next layer is metadata layer, based on the Resource Description Framework (RDF) that is essentially a simple relational data model, directed labelled graph, that can be serialized with a variety of syntaxes, such as RDF-XML, Turtle, or JSON-LD. RDF is used, for example, as the basis for the web ontology language OWL and logic standards for reasoning on the web based on Artificial Intelligence. The RDF model includes basic data types and can be used to define new ones for virtually any purpose.

In a typical Cultural Heritage application on the Web, data from legacy databases are transformed into a Linked Data knowledge graph [50]. It includes not only the (meta)data, but also ontologies that define the data modes used (e.g., Dublin Core or CIDOC CRM) and the concepts used is describing the contents (e.g., places, persons, times, and subject matter). The linked data is published using Linked Data principles and best practices of the W3C [51] in a SPARQL endpoint [] that facilitates re-use of data in research and application development [52].

A key challenge addressed by the Linked Data approach is interoperability of re-use of distributed heterogeneous datasets. The model is designed as the semantic, i.e., machine "understandable", basis for any content on the WWW by the W3C consortium leading the development of the WWW infrastructure. This is an important aspect in the European Museum Cloud that should be able to deal with heterogeneous dataset from countries and cultures.

Literature

A cloud directed to professionals working on Cultural Heritage subjects will need also to manage literature, i.e., either *published papers* or so-called *grey literature* (materials and research produced outside of the traditional commercial or academic publishing and distribution channels). In both cases the standard is now pdf. A European Collaborative Cloud for Cultural Heritage should therefore include pdf as one of the media standard data types.

Bibliographies

Compiling annotated bibliographies is a basic (often initial) action in multiple activities concerning Cultural Heritage. A system supporting Cultural Heritage professionals should support recollection and management of bibliographic references. A data model for bibliographies should be provided on the European Collaborative Cloud for Cultural Heritage (following available standards) and specific tools should be provided to support the creation of specific bibliographies.

2.1.3 Data quality and validation methodologies

Two important characteristics of visual data (and other types of data) are *provenance* and *integrity*. **Provenance** means that the person/institution who produced a specific digital asset should be known (info included in metadata scheme). **Integrity** means that the results of a digitization should not be changed, we should be able to assume that a given digital representation is not transformed/changed/altered (e.g., deep fake technologies); or, in case this happens, it should be properly described as paradata attached to the specific digital assets. Modifications may have multiple origins (cropping or skewing an image, changing colour gamut, cutting part of a 3D model or applying smoothing filters / filling holes, etc.) and are very hard to be detected if they are not properly specified in the data which describe the digital asset (even if AI solutions could improve our capabilities to detect manipulations).
Focusing on professional applications and on data reuse, the European Collaborative Cloud for Cultural Heritage should be designed to support by design these features (provenance & integrity).
The data model of the Cloud should allow to create and manage links interconnecting an original digital asset to all the derived assets, to allow users: first, to detect the origin of each derived asset; second, to be able to check the quality loss of a specific derived asset wrt the original model.

Moreover, data can come at **different levels of quality**, both due to the source (professionals vs. amateurs) or to the digitization process adopted and the planned use (data acquired at full quality vs. data acquired in a partial way or at low resolution following the purposes of a specific application). This issue would become even more pressing when a common cloud would be open to users having different qualifications and objectives (e.g., the very high-resolution images produced for a conservation project and the photographs used for a web presentation). We could already experience the different level of quality of the 3D models published so far on SketchFab.

The value and strategic impact of a European Collaborative Cloud for Cultural Heritage would be dependent by the clear specification of the level of quality and fidelity of any representation or data stored. If we want to base future studies on digital clones, it is mandatory to have some degree of confidence/trust of the digital clone, up to what extent it could be considered a correct representation of the real artefact.
Thus, this justifies the importance of ***documenting, assessing and validating data quality*** and the central role of related data models and instruments which could be included in the Cloud for supporting these features.

How to ingest data onto the platform, from the institution's point of view.
When preparing data for ingestion, Cultural Heritage institutions must follow specific guidelines. Part of this is adding metadata to data types such as objects and articles (or aligning their own existing metadata with the guidelines). The guidelines come with an extensive manual, to make sure that also those with less expertise can deliver the desired level of quality. The platform offers the ingestion tools to import the prepared data into the system.
The metadata will be the glue binding all the content on the platform together (and this is also a solution for the findability/availability of lesser-known data types, see below). There will be a thesaurus (or more). The guidelines will include one preferred notation of dates, locations etc. From the end user point of view, this activity is crucial; we have evidence of users experiencing problems after ingesting non-aligned content. This seriously affects user satisfaction and search query effectiveness. It is better to perform excellently within a smaller body of content than it is to do everything only half in a large environment.
This is why we need the commitment and participation of Cultural Heritage institutions. The platform cannot and will not accept all type or subtype of data. Voicing the museum's needs, the topic of data should be approached in a practical manner instead of striving for an academic completeness. Let's

take the Pareto principle to select a number of supported data types per sector (such as pdf for literature).

Data in virtual environments is usually presented as realistic and plausible 3D (or nD) information that can be interactively inspected by users and experts. However, acquired data is noisy for multiple reasons, including measurement and processing-based errors, and this uncertainty is largely ignored by present applications. The consequence is that most immersive (and non-immersive) visualizations nowadays are fundamentally biased and can be even unreliable. As an example, let us consider the digitization of a monument based on, e.g., a time-of-flight scanner. Most regions will have a dense sampling of data points, resulting in quite small errors in the final reconstructed shape. However, some other regions will probably suffer from lack of data, poor sampling, and inaccurate reconstruction. Final errors will affect both the geometric shape and its appearance attributes in a local way. Within most small regions of the inspected model, precise data will coexist with inaccurate geometric and attribute information. As hiding data errors can extremely hinder data comprehension in many cases, 3D models in applications requiring data comprehension should include geometric data and attributes plus local fidelity information. Understanding models created from acquired data requires a joint representation and visualization of data and data uncertainty. Quality representations are a key requirement if digital Cultural Heritage models are going to be used by experts for analyzing data, understanding the relations between different parts and taking measurements on digital models.

Reconstruction models typically rely on geometric prior assumptions, formulated independently of the input data. The most common assumption for reconstruction is spatial smoothness [44]. These types of approaches are now very well understood, and modern algorithms can for example fit minimal surfaces to noisy data in a globally optimal fashion [44], leading to very strong results. However, final 3D models do not contain error bounds on the reconstructed data. Moreover, simple smoothness assumptions limit the ability to handle larger holes in the data, as missing data is usually replaced by too smooth membranes.

Computing precise local fidelity bounds in huge reconstructed 3D models is still an open problem with interesting research challenges. While several works and methodologies have been proposed for the evaluation of the accuracy of active 3D scanning systems [45] and related bibliography), the evaluation of the final 3D model fidelity is still a neglected topic. A key document for understanding the importance of representing the uncertainty of measurements is [46]. It includes precise definitions of measurement uncertainty, the measure and specification, standard and expanded uncertainties, and measurement procedures.

Understanding huge amounts of acquired data requires a joint representation and visualization of data and data uncertainty. Models in 3D applications requiring data comprehension should, therefore, include geometric data plus attributes plus local fidelity information. This is clear by a reductio ad absurdum argument: having no information on data uncertainty will lead to wrong interpretations, wrong model measurements and, in most cases, lack of data comprehension. Let us first imagine a cultural heritage application to inspect the state of conservation of an ancient monument, with periodic acquisitions of its shape. Unless special care is taken during the overall acquisition and processing phases, it may occur that errors in surface data 3D positions are of the same order of magnitude as the erosion changes in the surface shape, making any expert analysis of the reconstructed models. In short, data can only be fully understood when information on data accuracy is available locally at any 3D point of the model.

The fidelity of a digital model is the (local) degree of approximation to the original object. It can be either geometric or related to any appearance attribute (colour, materials, etc.). Lack of fidelity can be a consequence of acquisition artefacts and dark regions, noise in the acquisition measurements, or approximating hypothesis during the subsequent geometry processing (model repair or simplification, for example). The fidelity of the final digital model is the aggregation of all individual error and noise sources. Moreover, we will use the term uncertainty to note specific fidelity bounds. Uncertainty is a measurable bound of the degree of approximation between the digital model and the physical object at any surface (or model) point.

Measurable 3D Models (MDMs in what follows, [46]) are digital models that explicitly encode local uncertainty bounds on their quality and fidelity. They encode geometric data plus attributes plus local uncertainty information. By encoding this model uncertainty information in a local way, MDMs include a measurable bound of the approximation between the digital model and the physical object at any surface point, by means of quantitative surface uncertainty maps, for instance.

By representing uncertainty, MDMs become fidelity aware. By encoding local model uncertainties, MDMs become able to compute uncertainty-based measurements (or measure uncertainties) as defined in [46]: measure uncertainty is a parameter, associated with the result of any measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand.

Local data uncertainty can be encoded in many different ways. Acquired models, however, are usually obtained from single acquisition sessions and they lack the series of observations that would be required to perform reliable statistical analysis. The consequence is that classical uncertainty measures like the standard uncertainty [46] cannot be considered in these acquired digital models. An alternative could be to use bounding intervals, in a way similar to expanded uncertainties [46].

Uncertainty/fidelity should also be properly visualized. The choice of visualization algorithms is critical as it will influence the user interaction and the final understanding. Anyway, model uncertainty information as stored in MDMs should not be confused with the final measure uncertainties during expert measurements.

The process of generating measurable 3D models must ensure that fidelity to the physical object/scene is preserved along the different steps of the acquisition and reconstruction process. Local fidelity information should be captured at the acquisition phase and properly transformed during geometry and attribute processing of the 3D models into uncertainty bounds. Algorithms based on Interval Arithmetic are especially well suited for this purpose, while statistical approaches like the one proposed by [47] could be useful in data coming from laser scans. In [47], the authors capture uncertainty by introducing a statistical representation that quantifies for each point in space the likelihood that a surface fitting the noisy point data passes through that sample point. In any case, the key rule is to avoid discarding information at any processing step.

Anyway, the main focus nowadays is on geometric fidelity/uncertainty. But, as already mentioned, Cultural Heritage models do not represent only the shape characteristics of an artwork. Representing at high fidelity the **appearance** (colour or surface reflection attributes of the artwork surfaces) is also a key factor in this domain. The evaluation or computation of the actual colour/surface reflection attribute fidelity is a quite open research domain, where methodologies are much less consolidated than in the shape-related case.

How to cope with data not including fidelity/uncertainty attributes

Nevertheless, digitized clones should come in the future with well-defined attributes measuring fidelity/uncertainty, for some time we will have to cope as well with models which do not include this information. Therefore, we will need some instruments to perform some sort of validation on these models. There are several possible interpretations of how assessment and validation should be implemented:

- *Data producer's reputation*: a first approach could be to link the potential quality of the data to the reputation of the institution (metadata/paradata assigned to each representation should include the institution/company in charge of data acquisition). But this is just a guess, not a very selective attribute.
- *Expert assessment*: peer-based evaluation of each specific data, based on the experience of the evaluator (subjective evaluation). This could be the better approach, but is costly and strictly depends on the experience of the evaluator.
- *Automatic assessment*: an automatic system could be designed, able to perform an objective evaluation of some specific attributes that overall allow to draw a qualification of the data. Some examples could be proposed for images (resolution, image sharpness) or 3D models (resolution, sample density measured as the average triangle size or inter-point distance, percentage of holes, verification of fidelity wrt. the original raw sampled data, etc.). We have approaches known in literature, but it is also a subject of potential research. So far mostly geometric or image-processing solutions have been proposed, while this domain could be also subject to innovation (adopting and customizing AI technology to cope with the quality assessment task).

The more, the better: all three options above could be part of a multi-variable evaluation profile.

2.1.4 Data preservation issues

Preservation is not just ensuring long lasting physical retention of the digital data (with continuous backup, replicated storage, use of permanent data memories, etc.), but also preventing data loss due to the obsolescence of data formats or data access applications.
Data formats are digital conventions, they evolve in time (very quickly in the case of visual technologies) and we already have memory and experience of the potential limited lifetime of specific encodings.
But we are proposing tools for a domain that studies the past, where professionals are used to work with documents and assets produced tens, hundreds or even thousands of years ago. Thus, a potential planned lifetime of a few years, which is acceptable for many CS applications, becomes totally not acceptable in the Cultural Heritage domain.
A second problem is with the applications which allow us to access, visualize and navigate the data. Custom programs could be easily discontinued, possibly because important components or libraries are no longer maintained (see the case of web-based applications based on the discontinued Adobe Flash).

Reuse of data is a key factor in helping us to preserve data. More a dataset is used, the more it will be maintained and, possibly, converted to new formats.

A European Collaborative Cloud for Cultural Heritage which wants to project itself to the future needs explicit preservation planning policies. The design of the system should include instruments and policies to ensure that data would be preserved, e.g., taking care of the status of each data format and providing exporting features to convert to other more resilient data types (preventing obsolescence).
Thus, some related requirements for a European Collaborative Cloud for Cultural Heritage are: it should be possible to detect all datasets belonging to a specific data type (to allow to easily recollect the ones that have to be converted, in the case of obsolescence of that data type); it should support instruments for automatic conversion from a data type to another data type (with impact on paradata: it should be possible to encode the eventual data degradation occurred in the migration process).

2.1.5 Digitization technologies: a consolidated domain or further research needed?

In the last decades, creating and dealing with digital replicas of artworks has emerged as an essential Cultural Heritage area that cannot be disconnected from real-world artefacts and monuments. We now have a large number of tools to create digital twins of these objects, but we still lack a usable, public, stakeholder-driven universe to view, understand, complement, share and retrieve the digital Cultural Heritage contents - thus, a *digital ecosystem*.

In fact, digital models will change in time. According to Eva Pietroni and Daniele Ferdani [53], "The digital object, if limited to the reproduction of a real object, becomes a replica... the concepts of uniqueness and authenticity need to be again pondered in light of the digital era. Indeed, real and virtual should be considered as a continuum, as they exchange information favouring new processes of interaction and critical thinking".

The digital Cultural Heritage universe includes visual data like images and panoramas, 3D representations, documents, terrain models, videos, or sound (see section 2.1). Three-dimensional representations can be either sampled and acquired from the real world, or modeled by design experts with usually lower and unknown fidelity, and they should include geometric information, local appearance/colour, local fidelity data (both geometric and appearance- related). This digital universe should feed the needs of experts when performing analytic investigation, also being able to generate holistic representations for the public including information like iconographic sources, literary sources, a narration, or architectural context. Tools over digital Cultural Heritage models should always be open, flexible, and adjustable, being usable and being defined from stakeholder requirements.

Besides models aimed at virtual restoration and stylistic intervention, virtual Reconstructions should always take into account the risk of mistaking for “truth” what is actually nothing more than a simulation [53]. Visitors should understand that, mainly in modeled 3D representations, what they observe may be the result of an interpretation (one out of many), not the "truth". In this context, and having in mind that the main objective of a museum, besides preservation, is education, 3D digital objects should be connected with information and related knowledge.

Of course, digital content should be designed in a way to fulfill community requirements, including galleries, libraries, archives, museums, and heritage scholars and communities. One of the digital Cultural Heritage-related difficulty is to devise a polyhedral set of tools being able to visualize, understand, complement, connect and even modify the digital universe from perspectives so diverse like those from museum visitors, students, experts on restoration, art experts, the research community, historians, or long-term archive experts.

A list of the main Cultural Heritage datasets that are publicly available, including OmniArt, WikiArt Paintings, BAM, IconArt, PrintArt, Arran, the Rijksmuseum dataset, Europeana, and the Web Gallery of Art, is presented in [54].

2.2 Coupling and integrating digital representations with metadata and semantics

In order to make Cultural Heritage content Findable, Accessible, Interoperable, and Re-usable, as required by the commonly agreed **FAIR** principles (https://www.go-fair.org/fair-principles/) metadata must be attached to the digital objects. This raises several questions to be addressed when designing/using the European Collaborative Cloud for Cultural Heritage:

1. What metadata models should be used for representing tangible and intangible Cultural Heritage contents in different application domains?
2. What vocabularies should be used to populate the metadata models when describing Cultural Heritage objects?
3. How to make the heterogeneous metadata based on different data models and vocabularies, and written in different languages, interoperable, so that the datasets can enrich each other?
4. How to make the data accessible via user interfaces?
5. How to implement the data creation pipelines for metadata?
6. How to publish the data models, vocabularies, metadata, data, so that they can be found and re-used easily by potential users?
7. How to maintain the metadata and data in a sustainable way, when the real world and its conceptualizations evolve in time?

These issues can be addressed on several levels, e.g., within one memory organization, national level, application domain level, or on a European level. In the latter case, the challenges of content interoperability between heterogeneous datasets for different countries are of central importance, as has been learned when developing the Europeana service. Below, we focus on this especially on this issue [55].

The process of making heterogeneous datasets interoperable with each other is known as data reconciliation. Data reconciliation must proceed at two different levels – syntactic and semantic – because datasets can be heterogeneous in two different ways. In linguistics, syntax relates to the structure of a sentence and semantics to the meaning of the individual words and phrases within it. Similarly in computer science, syntax refers to the manner in which data is structured, and semantics refers to the significance of the individual elements within those structures and their relations. One set of tools is needed for reconciling datasets which differ syntactically (i.e., are structured differently). A different set of tools is needed for reconciling data which differs semantically (i.e., which express the same meanings in different ways).

2.2.1 Syntactic interoperability and data cleaning

On the syntactic level, the same data can be structured in many different ways. For example, an inventory of correspondence consisting of precisely the same data can be presented in tabular formats such as CSV (comma separated values), or as JSON (JavaScript Object Notation) objects [56], or as RDF graphs (Resource Description Framework) [57]. In addition, data values, such as dates, person names, and numeric values may be represented in different forms: the same Gregorian date, for instance, can be represented as *yyyy-mm-dd*, as *dd-mm-yyyy*, and in many other configurations as well. In order to be made interoperable, these differing data structures and representations need to be transformed into a common format: for instance, when publishing Linked Data, everything needs to be transformed into RDF.

A task closely related to syntactic reconciliation is *data cleaning*, that is, the removal from the data of typing errors and irregular formatting. Syntactic transformations are in many cases technically fairly straightforward to do but may be tedious. There are many tools available to facilitate data cleaning and transformations, such as OpenRefine [58] and Karma [59].

2.2.2 Semantic interoperability

More formidable are the challenges encountered in data reconciliation at the semantic level. Reconciliation at this level is needed since data can mean different things even when expressed in similar or identical ways. For example, the date 1 January 1600 means different things depending on whether the Julian or Gregorian is used. Similarly, the name ‘Neustadt’ could refer to over twenty different places in Germany alone and a dozen more elsewhere, not to mention to innumerable people with that surname. On the other hand, completely different synonymous names can also refer to the same entity: ‘Leiden’, ‘Leyden’, and ‘Lugdunum Batavorum’ all refer to the same place in the 16th century.

Further difficulties arise on a higher structural level, where incompatible ways of representing knowledge frequently occur: for instance, if the creator of a text is indicated as an ‘author’ in one dataset and a ‘writer’ in another, then ‘writers’ are not found when looking for ‘authors’, or ‘authors’ when looking for ‘writers’, both of which lower recall. In many cases, the semantic content in one dataset cannot be represented by using the knowledge structures of another dataset. In such cases, a more fundamental underlying knowledge representation scheme is needed for representing both datasets and their relations in common terms. To represent the different expressions and manifestations of the Bible in print or in other media formats, for example, a deeper notion of the underlying idea of the Bible as an immaterial work and its physical representations is required. The same applies to the individual letter, not to mention people and places.

2.2.3 Major approaches

This section considers data reconciliation from the perspective of integrating cultural heritage linked data contents that are represented using different kinds of metadata models [60]. In this context, two major approaches are in use for reconciling (meta)data. First, within the Dublin Core framework [61], different document-based schemas can be harmonized by using the Dumb-Down Principle. The idea is to map metadata elements onto each other within a hierarchy. For example, if the ‘author’ and the ‘writer’ of a text are represented as sub-properties of a more general property ‘creator’, the machine can understand the relationship between the different kinds of creators. Similar hierarchies can be established for other classes of concepts. This allows queries to be expanded by moving up the hierarchy: searching for ‘creator’ returns more results than either ‘writer’ or ‘author’. Alternatively, a fundamental underlying ontology describing the domain of discourse can be modeled and different metadata schemes transformed into it. The term ’ontology’ here refers to ontologies as formal, shared, structured models of data used in Computer Science for representing knowledge, not to ontology as a branch of Philosophy studying the nature of being, existence, and reality [62].

The best-known examples of this approach are the CIDOC Conceptual Reference Model [63] for cultural heritage museum data, and the Functional Requirements for Bibliographic Records (FRBR) (and the related conceptual models FRAD and FRSAD) for library data [64] that were recently consolidated as the IFLA Library Reference Model [65]. CIDOC CRM and the FR-models are being combined into a Conceptual Model for Bibliographic Information in Object-Oriented Formalism (FRBRoo) [66]. In contrast to the document-centric approach of Dublin Core, the conceptual reference models above are event-centric in nature: within them, the world of discourse is described in terms of events and their related constituents – especially participants, place, and time – in a manner highly suitable to modeling various forms of tangible and intangible objects and phenomena in the Cultural Heritage sector.

2.2.4 Semantic disambiguation and entity linking

A recurring basic problem in data reconciliation is how to map literal, i.e., textual expressions in (meta)data, such as names of persons and places, onto their corresponding unique meanings, defined in a reference registry (domain ontology). This process is called entity linking (NEL) [67] and ensures, e.g., that the same persons or places mentioned in the metadata can be accurately identified and the data interlinked [68]. A key challenge in NEL is semantic disambiguation, that is, the task of selecting the correct referent from multiple possible choices.

When dealing with historical Cultural Heritage materials, we often lack the contextual data needed to disambiguate mentioned entities with certainty or even to identify a referent candidate tentatively. This commonly leads to two equally undeniable outcomes. What is needed in order to avoid these opposing difficulties are methods for dealing with uncertain links between named entities.

Depending on the desired level of precision and recall in NEL, fully automatic semantic disambiguation and linking may not be possible. Fortunately, it is often possible to identify problematic instances and to develop semi-automatic tools where difficult cases can be resolved with the help of a human expert.

2.2.5 Supporting FAIR Principles

To support the FAIR principles on a European level, developing a shared Cultural Heritage data infrastructure is needed. Instead of developing organization and national custom metadata models, ontologies, and other solutions, more collaborations and agreements are needed between the stakeholders in different countries, even if this complicates local processes for data production and publishing in local data silos. This price should be paid, if the benefits and savings on a European level are bigger.

A shared data infrastructure for a European Collaborative Cloud for Cultural Heritage would need several components, including the ones below:

- Ontology/vocabulary services with APIs for re-using them.
- Data services for publishing data models and datasets for re-use and application development.
- Software tools for aggregating, producing, publishing, and analyzing datasets.
- Best practices for cataloguing, publishing and using Cultural Heritage data.
- Human infrastructure for using the data infrastructure.

The European Collaborative Cloud for Cultural Heritage could be an instrument to facilitate this. In general, it is wiser and cheaper to support the local content creation to prevent interoperability problems than trying to fix the data problems afterwards when the damage has already been done. Albert Einstein has said: “Intellectuals solve problems but geniuses prevent them”. This wisdom applies here, too [69].

2.3 Interaction with Cultural Heritage data/content

This section presents approaches and technologies needed for interacting with Cultural Heritage data.

2.3.1 Search and information retrieval

Cultural heritage (CH) services on the Web publish typically collections of objects in museums, libraries, and archives. Interaction based on search and information retrieval is a most usual way of interaction with such systems. A typical use scenario is: 1) first to find an object or, more generally, a set of objects of interest and 2) investigate the search results by browsing for more information about the object, recommended related objects, or about contextual data. It may also be possible to analyze or visualize the results.

The simplest search interface is illustrated by the Google search page, where only on a text field used for user input. In a more advanced search, the user may search simultaneously separate metadata fields describing the objects, such as object type, location of manufacture, time of production, etc. A more and more used search paradigm is faceted search [84] where the objects are classified along orthogonal facets, such as the metadata fields above, and the user searches the data by making category selections on the facets one after another in free order. After each selection, say object="vase", a new results set is computed, and a predictive count of results for each possible next selection. In this way, the user is directed towards feasible refinements of searching and "no hits" results can be avoided. Faceted search can be integrated easily with data analytic tools: after each selection the result set can be, e.g., be visualized on maps or on a timeline, and statistics be counted along the facet categories. Arguably, there is a trend in Cultural Heritage systems from search and browsing-based data exploration systems towards systems supporting also data analysis, and finally towards knowledge discovery and Artificial Intelligence-based systems, where the computer may actively search or suggest the users new research questions, solve them, and even explain solutions. [85]

More and more data is accumulated on Cultural Heritage problems, made accessible to professionals, the more professionals will adopt **data analytics** and **data visualization** technologies, to enable new research paradigms and to foster the capability of inferring knowledge from available data.

2.3.2 Visualization and inspection of Cultural Heritage data

We presented the many media we should be able to manage in the Cultural Heritage domain in section 2.1. Each of those media requires a specific visualization component. There are some issues which make the implementation of interactive visualization tools quite complex, but possible: managing **data size**; providing **consistent GUI** for visualization; enabling **multi-platform delivery** (web-based).

Managing the **data complexity** is a must in Cultural Heritage visualization. The quality of current digitization technologies allows to produce huge models (giga-pixel images; extremely dense 3D models; etc.). More than 20 years of research on complex data management has produced an arsenal of consolidated technologies which allow to represent at multiple level of detail or in multi-resolution the original data, to compress those data, to apply view-dependent rendering modalities based on progressive transmission, and finally endorse GPU-enable rendering (whether available). All these technologies allow to render in real time highly complex dataset even on low performance devices. Adopting these technologies is a must for ensuring interactive performances of a data inspection tool.

We already mentioned the need of providing **consistent GUIs** while inspecting different media (a very simple example is the zooming action, which should be implemented using the same approach either we are inspecting an image or a 3D model; others are trackball manipulation, panning, selection of point of interest, etc.). This issue was highlighted by some recent works of the YALE CS Dept [70] leading to the definition of the CHER-Ob system, an open-source tool providing a single intuitive interface consistent with multiple media types (2D images, hyperspectral images, volume data encoded using the DICOM standard and triangulated 3D models, with or w/out textures) and implementing just Cultural Heritage needs.
In the multiple media context of Cultural Heritage application, this issue should be taken into account, thus leading to a common design of the GUI of the multiple visualization tools.

Another issue is to provide a **consistent approach** also on **different platforms** (PCs, tablets and smartphones). Here a turning point has been endorsing web-based technologies, able to render on any device any data, assuming a webGL-enabled browser is implemented on each device.

Therefore, implementing **efficient visualization** tools is possible but those constraints should be taken into account. There are already several examples of visualization tools, either commercial resources (usually, designed for other application domains) or open-source solutions (some of them results of previous EC projects). Just a few examples:

- Sketchfab (https://sketchfab.com/categories/cultural-heritage-history) is a commercial market-leader visualizer (supporting only 3D data).
- An example of open-source platform for multi-platform interactive visualization is the Visual Media Service [71] (http://visual.ariadne-infrastructure.eu/), implemented using 3DHOP technology [72] and the Nexus multi-resolution representation [75] (http://vcg.isti.cnr.it/nexus/).

To explain further which are the features needed for visual data inspection in Cultural Heritage, we list here some specific features:

- **Navigation**: support of easy navigation *around* the digital artefact (the usual approach for 3D objects is based on the *trackball* interface metaphor, allowing to rotate around, zooming and panning over a object) or *inside* a 3D space (navigation inside complex scenes, e.g. a virtual museum, that requires a different interaction modality and should be designed to ensure easy access of different classes of users).
- **Selective rendering modes**: the tool should provide different options to drive the visualization, e.g., shaded geometry, setup of light settings and change of incident light direction, on/off colour/texture mapping, shadows, transparency, etc.
- **Measuring**: support of measuring features (Euclidean or over-the-surface distances among selected points, surfaces, volumes), computed on the flight following the requests of the operator; features and metrics to support similarity analysis (for the sake of comparison, classification and interpretation).
- **Cut-through sections**: possibility to virtually cut the geometry following selected cutting planes, useful to investigate the shape of the object or the relation between interior and exterior shells.
- **Annotations**: users may need to link comments or knowledge to the digital clone, in forms of annotations associated to specific locations of the artwork surface (see subsection 2.3.5).
- **Mixed-media rendering**: the possibility to render multiple media in the same visualization context [43] might be extremely useful to analyze an artwork, either because we might need to cross-compare images and a 3D model [74] or because we might need to analyze jointly the exterior (a 3D scanned shell) and the interior (data from a TC scanner) of an artwork [76].
- **Snapshots**: during interactive navigation, the visualization tool should support the acquisition of snapshots (images) of the current visualization canvas (active window or subsection), to document a specific view of interest in the current reasoning or study process. A similar feature is the possibility to save the **current view specs**, to be able to create the same view in a future time (this requires to save all parameters, view point, zooming factor, shading used, etc. in a machine-readable format).
- **Uncertainty/fidelity visualization**: the visualization tool should support the user in inspecting or visualizing the uncertainty/fidelity of each digital representation, either globally or related to any location of the model; it should return to the user either numeric data or support a visualization mode able to present the fidelity (e.g., using a false colour ramp while shading the surface).
- **Temporal representations** (also known as 4D): in the case of models encoding the evolution on time of artistic or architectural objects, the tool should provide features and an interface enabling the visual analysis of the 3D model over time (not just a time slider, but instruments to facilitate grasping the evolution on time of the represented phenomena, helping in spotting and measuring the changes).

Most of these features are already quite consolidated (thus, lead to IA actions), while others require some further research effort (thus they might be funded in the framework of RIA actions). Examples of the latter are: navigation of complex scenes (easy to use navigation metaphors, effective on multiple platforms and different interaction devices); mixed media rendering.

From the user point of view, it will be extremely dissatisfying if features like measuring, navigation, annotations function in only part of the content. 100% is impossible, but it is better to do excellently in a smaller environment than it is to do everything only half in a larger body of content.

2.3.3 Sharing and publishing Cultural Heritage data

The availability of technologies for sharing data and for publishing them on the web is a key element for a European Collaborative Cloud for Cultural Heritage and for implementing the vision proposed in this report. The Cloud should offer to Cultural Heritage professionals (and even common citizens) very easy-to-use instruments enabling efficient web-based **data sharing**.
But data is often the result of effort and money dedicated by institutions or professionals, thus often there is some level of confidentiality in their use. Therefore, providing controlled **access** and **IPR** protection is also a key element for a wide acceptance of a data sharing approach.

Technologies for sharing data or publishing on the web
Data sharing is nowadays performed using web technologies and offering front end interfaces for data ingestion and data access. The old mode (based on file transfer) is now totally replaced by a much more effective approach, where the search is performed on a web page and the resulting digital objects are immediately accessible and visualized using (enhanced) web technologies.

Several efforts have been made to offer *easy sharing* of *3D data*. A recent paper [32] by Erik Champion and Hafizur Rahaman presented a critical review of some existing online 3D repositories, focusing on their goals, the offered hosting features, and the capabilities of their 3D viewers. This work and its results have already been mentioned and discussed in subsection 1.3.5. We recall here some conclusions: many approaches have been proposed, but they are disconnected, interoperability is still not granted; most of them lack a proper management of *data provenance* and *metadata*; most of them support only the interactive visualization and other important Cultural Heritage-related features are not provided (annotations, measures, check of similarity, links to other sources of information or data, etc.).
Two representatives of tools supporting web-based publication and sharing are: on the commercial side, the **Sketchfab** service (https://sketchfab.com/); on the open-source academic side, the **3DHOP** platform (https://www.3dhop.net/index.php) or the related **Visual Media Service** (https://visual.ariadne-infrastructure.eu/) tool.
Recent experiences (pioneered by Sketchfab and Visual Media Service) demonstrate that the simplicity of the ingestion phase is a key element in the success of a data sharing tool. Ingestion should be as much as possible easy for the data contributors; ingested data should be automatically checked, possibly enriched and transformed in formats enabling efficient transmission and visualization in a web context.

Moreover, Cultural Heritage data are inherently multi-media (see section 2.2). Supporting publishing and sharing Cultural Heritage data should include all types of data, using common and congruent interfaces both at the ingestion and at the data access/visualization stages.
The **Visual Media Service** is the only platform so far offering management of several types of data (2D images, collections of 2D images, RTI images, 3D models) with a common and uniform interface.

IPR issues

The accumulation of Cultural Heritage data is often the result of effort and money dedicated by institutions or professionals, since digitization is still a costly process (when performed according to full quality standards and procedures). Moreover, many digitization actions are performed in the framework of an activity focused on a specific Cultural Heritage subject(s), following research or conservation/restoration interests. In these cases, data are part of a current research/professional activity and thus there is a confidentiality to be respected in their use. Professionals could be strongly against the idea of publishing data on a shared platform if those data are the current subject of an open activity.

But demanding data ingestion at the very end of a project (the final phase, where data are cleaned and decisions on what could be shared and published are taken) is also a risky approach, leading in many cases to the lack of specific decisions or to partial publishing, abandoning most of the data on inaccessible private repositories (which would increase substantially the risk of data loss).

It is therefore mandatory to provide *controlled access* to the data. Users (data owners) should be able to define the specific data access policy for each ingested data token (accessible just to the owner, or to a research group, or open to the public). Moreover, the usability of each data (e.g., following the Creative Commons licenses, https://en.wikipedia.org/wiki/Creative_Commons_license) should be properly specified, to protect IPR.

An important factor in convincing users to contribute content to a cloud is also to support easy accountability of the creator of each data, at all the different levels we envision for the European Collaborative Cloud for Cultural Heritage (from the digital clone, i.e., a specific 2D or 3D representation of an artwork; until the extreme of data representing the digital continuum of an artwork). The Cultural Heritage community requires proper acknowledgement of who and how is using materials they have created and curated. A platform which will explicitly manage ownership, rights of use and effective re-use of the data contained will hopefully increase the willingness of professionals/scholars to contribute to the archive. It will also stimulate and support the endorsement of improved evaluation policies in SSH (adding indicators related to data production and data reuse to the usual publication-related indicators).

To conclude, the source of the ingested content will be tagged as the owner of the data and will remain responsible for it, as per the European Collaborative Cloud for Cultural Heritage platform's terms & conditions. The source will also be responsible for defining if all, or part of, the content could be made accessible in the public domain, when this will occur (could be delayed wrt. the date of ingestion, to fulfill temporary confidentiality) and, finally, under which CCH license.

2.3.4 Passive presentations – Video & Computer Animation

Videos are a common instrument used in museums or for educational use. Video technologies are an ICT domain with a very solid market, both considering video production systems (video editing) or systems for producing computer animations from 3D models or 3D scenes. Those tools evolved constantly in the last 30 years and nowadays we have plenty of video editing or computer animation commercial tools. YouTube™ consolidated a leadership role in offering archival and streaming capabilities.

Due to the mature status of the industrial products, a new European Collaborative Cloud for Cultural Heritage could endorse those instruments without duplicating efforts.

2.3.5 Designing interactive installations

Interactive installations play an important role for museums, since they allow visitors to play with digital clones, to discover them and to present a corpus of information on the subject via clickable links or navigation options.
An important characteristic of a platform for creating interactive content is to allow users to design content which could be presented to museum visitors in the museum (with an interactive kiosk, a PC or a tablet), and being published on the web to grant broader access. Having the possibility of designing content which could be visualized both locally and on the web is important to maximize the fruition/impact, reducing the development costs.
A platform for designing interactive content thus should support web-based visualization and interaction, should support the use of multiple media, should provide easy instruments for customizing the look and feel of the application interface, and adapt it to the graphical style of other museum content.

The term ***storytelling*** is used for a more sophisticated approach in telling the story of an artwork, using narratives [77]. Several experiences have focused on the design of authoring systems enabling storytelling approaches in the creation of MM content.

2.3.6 Immersive technologies (VR/AR/XR)

Immersive visualization technologies are becoming a new way to access, visualize and understand Cultural Heritage objects and monuments. Present technologies include stereo vision and real-time user tracking, connecting the body movements of the user(s) to the virtual camera in order to create a perception of being in the virtual environment. Sound and tactile (haptic) feelings are also possible, although tactile feedback is still rare.

Immersive experiences should be believable (feeling like being actually there), interactive (the system should react to user movements in real time, to convey the feeling that we are inspecting what we intend to), immersive (giving visual and sensorial believability) explorable, allowing users to walk around the VR environment [78].

We should distinguish among virtual reality, augmented reality and mixed reality (for a discussion among these three alternatives, see for instance [79]):

- Virtual reality systems (VR) aim at creating the sensation, in the visitor or user, of "being there". VR transports users to a virtual environment without any or little possibility of directly interacting with their immediate physical surroundings [79]. VR systems can be either projection-based or fully immersive. In the first case (CAVE systems, PowerWalls etc.), users must be in a specific room with stereo screens while wearing special glasses (active or passive) that allow them to observe both the virtual environment but also their colleagues in the experience. Fully immersive VR systems are based on headsets (like Oculus Quest and others). They give a strong immersive feeling, but the VR experience is individual, interaction with the rest of the people in the group being almost completely lost. On the other hand, some people may experience dizziness.
- Augmented reality systems (AV) merge virtual environments with events and elements from the real-world. While VR is basically disconnecting users from their perception of the real world, this is somewhat achieved by AV achieves since live real scenes are seen (or streamed). They have an interesting potential use for in-situ enhancing of the visitor's experience, by showing, for instance, virtual reconstructions of destroyed parts of the monuments on top of the present remains, or by offering the possibility of switching among different reconstruction hypotheses also being able to show the evolution of the monument along the centuries.

Unfortunately, present technologies (Hololens and others) still lack optimal solutions to the problems related with 1) visual merging between the real world and virtual scenes and 2) real-time registration of both worlds.

- Mixed reality (XR) solutions work by enhancing the real monuments with virtual information. They are beginning to be widely used because of their robustness and quality. One of the examples could be the use of video mapping for the recreation of the original frescoes (see for instance the Pantocrator project (https://pantocrator.cat/projectes/) on the on-site reproduction of the Romanesque paintings in the main apse of Sant Climent de Taüll church, with original paintings from the 12th century that are now preserved in the National Art Museum of Catalonia in Barcelona). Video mappings can be certainly informative for visitors in general.

Immersive technologies have a significant potential in inspecting and understanding Cultural Heritage objects and monuments, although really usable platforms, designs and implementations are still not commonly available to the Cultural Heritage community [80] with some exceptions as the already mentioned video mappings. Semi-immersive VR systems open to general visitors have been used in several exhibitions (see for instance [81]), and in some cases of monuments having large open spaces, VR platforms can also be helpful to experts in their daily design work [82] to help them to understand the real dimensions of the spaces.

Virtual reality settings provide a strong experience of presence in non-existent (or distant, remote or disappeared) environments, with the aim of providing the same experience that users could have at the real site. Anyway, and besides their future potential, AR technologies are still immature while VR systems are either quite complex (in the case of VR-based settings) or lack ergonomic and real group experiences (in the case of head-mounted displays, also inducing motion sickness in some users). VR systems still have a way ahead to evolve to ergonomic sunglass-like elements with full capabilities and no cables.

VR systems could be suitable for inspecting specific spaces and monuments, but they are not a general-purpose solution. The convenience of using immersive technologies instead of standard interactive settings (see subsection 2.3.4) should be carefully discussed from the user point of view, focusing on the real needs and usability and not on simply showing futuristic technologies. It may be that in many cases, non-immersive systems mixing 3d models with good visual interfaces plus metadata and videos (see for instance [82]) become good alternatives. But still, immersive interaction will become a key tool for Cultural Heritage digital material in the next future. An important avenue of future research will be on user-centred VR, to design specific immersive tools and systems driven by the requirements and needs of Cultural Heritage experts, museum visitors and stakeholders.

2.3.7 Annotations

Annotations are an important component or feature in interactive systems.
We can consider annotations as a key feature of the system, allowing the expert to insert knowledge by linking it to the digital representation of an artwork. Annotation can therefore link any type of *information* (a short text, a longer report encoded with a pdf file, a tag or the value of an attribute, an image or a video, a recorded sound or a grabbed spoken stream) to a specific *location* or surface *subset* of the artworks (thus geo-referencing the information to a specific portion of the object).
Interaction plays an important role while defining annotations, since it requires a structured dialog with the user. The easiest form of geo-location is a *point-based* annotation, where the user has only to select a point over the external surface of the artwork. More complicated selections are the ones

based on: *subsets* (only the hand or the head of a statue), *polylines* (a set of interconnected lines) or *regions* (a closed polygonal portion of the surface).
Annotations are also important to introduce content that can be selectively disclosed to the user while he is navigating the artwork and related data with an interactive installation.

Annotations can be implemented following different approaches (see the review provided in [73]).

We highlighted previously the importance of providing multiple types of media to represent different features of a Cultural Heritage artwork. Thus, it is important to organize and implement annotations in a way that would allow to propagate the annotations defined over a single media to any other media. An example of a system designed following this approach is the AIOLI restoration documentation system developed by CNRS [74]. AIOLI manages point-based 3D models, RGB images and RTI images. An annotation added to an image can be projected to the 3D model and from this model to any other image sampling the same area. Automatically inheriting annotations on any media supported is an important plus of this system.

2.4 Potential uses for museums or Cultural Heritage institutions

In this section we list several activities of interest for Cultural Heritage communities and related technical methodologies and tools. We use the plural, Cultural Heritage communities, to underline that the Cultural Heritage domain is not a single body, it includes many different professional and non-professional activities, and thus multiple related needs.
Most of these needs are not satisfied by commercial solutions and thus are potential candidates for the design of specific tools to be plugged on the European Collaborative Cloud for Cultural Heritage (see the description of added tools in section 3.3, which derives from the suggested list of potential uses presented here).

All those tools will need to manage a subset of, or even all, the different media described in the previous subsections. Therefore, the data models of the European Collaborative Cloud for Cultural Heritage will be the basis to build all possible higher-level tools and instruments. The vision is thus of a European Collaborative Cloud for Cultural Heritage offering access to *raw data*, but also (and probably more important) to an arsenal of *digital instruments and tools*. Assembling data with digital tools providing functionalities for the analysis, investigation, documentation of the insight process could revolutionize the way Cultural Heritage-related activities are performed and consolidate the impact of digital technologies in this context.

The scope of the basic European Collaborative Cloud for Cultural Heritage platform is therefore to offer: the basic layer for data management and access, very few priority tools (which have to be intended also as examples of the applications which could be hosted by the Cloud); and an infrastructure (low and mid-level libraries, API) which will offer the opportunity for others to develop tools, apps, plugins etc. to be hosted on the Cloud (see Part 3 for a detailed presentation of these themes). Therefore, a **policy** and an **organizational model** are needed to ensure the safe evolution of the Cloud content (i.e., increasing number and scope of the applications served), the cooperation among different consortia and projects, and finally the preservation of data and content (see section 3.4). The **preservation** task in this case also includes the need of governing the maintenance of the different software components. Outside suppliers may include Cultural Heritage institutions, scholars, research institutions, commercial corporations, etc. An organizational and management policy will be needed, to govern the joint work of many different players, and to define a proper **business model** (who pays for maintenance and extensions of the European Collaborative Cloud for Cultural Heritage?).

2.4.1 Data Interface (ingestion, search & retrieval)

This tool will constitute an interface to the underlying data repository. It will support instruments for easy data ingestion in the repository (data and associated metadata); the GUI will be designed to reduce the complexity and human effort in the specification of metadata (CIDOC-CRM).
It will also provide an interface for issuing searches over the repository and for presenting the results of those queries; in the case of data which have a visual representation, it will support native visualization and inspection with web browsers specific for each data type, but having compatible interfaces.

2.4.2 Tools for metadata creation and enrichment

Metadata ingestion/enrichment has been so far mostly a manual phase, requiring considerable resources to be performed on massive digitization. New AI technologies could innovate considerably, by building links, enriching metadata, supporting the construction of semantic networks for Linked Data in a semi-automatic manner subject to human experts' supervision and validation.

CH collections on the Web are mostly metadata [91] about the actual objects and content in memory organizations and in the real world (e.g., Cultural Heritage sites, buildings, natural history places, etc.). Even with born-digital objects on the Web, metadata is needed in order to make the objects findable, accessible, interoperable, and reusable according to the FAIR principles.

Producing FAIR metadata is a key challenge if we want to publish and enrich European Cultural Heritage on the Web that originates from different countries, diverse Cultural Heritage domains, and is represented by mutually incompatible metadata models, written in different languages. This is a lesson learned, for example, in the massive Europeana.eu initiative, and in more domain-specific projects, such as ARIADNEplus for archaeology.

To mitigate the problems, the European Collaborative Cloud for Cultural Heritage could provide services related to the following areas of metadata production for the Cultural Heritage organizations:

- **Shared metadata infrastructures**. Pan-European shared data infrastructures for representing Cultural Heritage metadata are needed. Without shared ontological views and standards for knowledge representation Cultural Heritage organizations create separate data silos of their own and miss the opportunities for sharing and enriching their contents with other organizations and on an European level.
- **Metadata extraction**. The challenge here is how to produce structured metadata from unstructured DH data in memory organization, such as speech, texts [92], images, and videos? Here various methods, tools, and resources for information/knowledge extraction are needed. An example of this is the AI-based BERT language model by Google widely used for extracting knowledge from textual data. Transcribus https://readcoop.eu/transkribus/ is a success story of an EU project for creating tools for extracting text from images of handwritten historical documents and manuscripts.
- **Metadata alignment.** How to link existing and new metadata in Cultural Heritage organizations to the shared metadata infrastructure? Here ontology services for making the use of the vocabularies used in the shared infrastructure easy and cost efficient are needed. Also tools and resources for aligning vocabularies are needed as in EU Vocabularies initiative https://op.europa.eu/en/web/eu-vocabularies/alignments. This is needed for the Cultural

Heritage domain, too. An example from the U.S. is the effort towards this goal produced by the Getty Research Institute with its linked open data vocabulary services for the museum domain. For libraries there are systems such as VIAF for authority files. Bioportal https://bioportal.bioontology.org/ is a vocabulary service of some 1000 shared vocabularies and ontologies in the biomedical domain hosted by the National Center for Biomedical Ontology.

- **Metadata enrichment by data linking and reasoning.** Once machine readable semantic metadata is available it can be enriched by linking it to related data by other organizations and by reasoning using methods of knowledge discovery. Again, the European Collaborative Cloud for Cultural Heritage could host needed tools, services, and best practices for this can be done.

Experiences on creating and using a national level data linked data infrastructure for Cultural Heritage are described in [93].

The needed data interfaces here provide possibilities for storing an downloads datasets and other resources, APIs for re-using the resources of the Cloud in the legacy systems of the memory organizations, in research projects, and in application development by the creative industries, and portal interfaces for finding and learning to use the European Collaborative Cloud for Cultural Heritage services and resources.

2.4.3 New approaches for 3D/2D digitization

Digitization technologies are now quite consolidated for a large subset of the possible Cultural Heritage targets. Accuracy and sampling density supported by currently available technologies are now adequate for many applications.
Nevertheless, there are still difficult situations and practical reasons that strongly motivate further research:

- New **AI-powered solutions** to improve the digitization process of tangible Cultural Heritage assets. Increasing the robustness and efficiency of the 3D digitization process, especially in the case of massive digitization (collections of artworks); increasing the accuracy and completeness of surface appearance acquisition and further mapping of complex reflectance data on digital surfaces; post-processing and cleaning of the models produced. Solutions to go beyond the visible spectrum.
- Designing methodologies and instruments for computing and encoding precise local fidelity bounds in reconstructed 3D models (producing measurable bounds of the approximation between the digital model and the physical object at any surface point). Future 3D models should encode geometric data plus attributes plus local uncertainty information. How to derive or compute global and local fidelity figures is the focus of this possible action.
- Digitization of physical characteristics and behaviour of **complex assembly** or **dynamic** objects. While shape and appearance have been the main focus of 3D digitization efforts, the task of devising practical technologies to extend the digital documentation to the capture of information that allows sound physical simulations of the digitized objects remains still unexplored. Stiffness and flexibility, mechanisms and moving parts, mass distribution and resistance, are all aspects that still remain undocumented and, once properly digitized, could allow better analysis and understanding through more truthful simulations of ancient relics like tissues, mechanisms and tools. Many potential subjects present some dynamic behaviour, examples are all the machines or apparatus developed by humans, from archaeological times (e.g., the Antikythera machine) till modern time (e.g., all the machines which constitute the heritage of XIX and XX cent.). Modeling a dynamic assembly is quite a costly effort and often requires disassembling the machine, which is often not possible. Specific digitization solutions

can be devised, including alternative approaches requiring a lower digitization cost but giving the possibility of experiencing the dynamism of a mechanism without requiring the full cost of building an accurate clone. Mixing media could be a solution to this task: while a 3D model could represent well the static essence of the mechanism, other media such as video, sound, CT scans could sample the other more dynamic or hidden characteristics; all these media should be presented to the user in the same common interactive presentation context.

2.4.4 AI-based methods for automatic data analysis and knowledge discovery

AI technology could bring a stunning contribution to Cultural Heritage-related activities. AI-based solutions have already demonstrated great impact in many intelligent, usually user-driven tasks.
A limiting factor for Cultural Heritage applications of AI has so far been the scarce availability of data related to Cultural Heritage activities, to support the learning phase; the European Collaborative Cloud for Cultural Heritage could overcome this limitation.
This action will experiment with the application of AI methodologies or enhanced methods to solve research questions concerning insight over the data, such as: categorization, segmentation, recognition, analysis, understanding. The goal will be to replace usual user-intensive activities with automatic processes (which should be explainable, and the results supervised by human experts).
There is a wide set of potential users: museums, industry, scholars, collectors, citizens science, etc.

2.4.5 Digital Clones for Museum Curators

This tool will allow Museum Curators to structure, encode, store and analyze all knowledge needed to support curation activities.
The data archival will be based on an accurate and high-resolution representation (2D, 3D, multiple). The system will allow the curator to store and interlink all needed knowledge. This will entail:

- Basic archive and museum catalogue info
- Bibliography of literature related to the artwork
- Conservation history (including previous restoration actions)
- Loan and travel history (for temporary expositions), including links to related digitization to fix the status of the artwork before and after the loan

Here the focus is not just in documenting and archiving multiple heterogeneous documents, but also identifying/storing/visualizing connections among those assets; the digital clone of the artwork could be the spatial index for structuring and presenting all this knowledge. A cooperative approach will be enabled.

Museums can also be interested in technologies able to:

- Perform virtual restoration over degraded documents (e.g., virtual restoration of manuscripts, removing degradation and increasing readability).
- Detect copies (media spreading) and fraudulent/non authorized uses of digital assets representing artworks of their own property (e.g., non authorisednon-authorized commercial uses), to protect IPR.
- Keeping a tight control over the integrity of the digital representations (detecting changes or modifications, keeping a link between each derived asset and the original digitized asset). In both cases, AI-based technologies could contribute to an innovative and improved management.

2.4.6 Digital Clones for Art Historians/Scholars

This tool will allow Art Historians/Scholars to structure, encode, store and analyze all knowledge needed to support the study of the artwork.
The data archival will be based on an accurate and high-resolution representation (2D, 3D, multiple). The system will allow the Art Historians/Scholars to store and interlink all needed knowledge. This will entail:

- Basic national catalogue info
- Bibliography of literature related to the artwork
- Results of scientific diagnostic investigation analysis
- Results of inspection and study (annotations over the entire or selected parts of the artwork), to allow to preserve all evidence found to support Art Historians/Scholars hypothesis and to document the study process

Here the focus is not just in documenting and archiving multiple heterogeneous documents, but also identifying/storing/visualizing connections among those assets; the digital clone of the artwork could be the spatial index for structuring and presenting all this knowledge. A cooperative approach will be enabled.

2.4.7 Digital Clones for Restoration

This tool will allow Curators and Restorers to structure, encode, store and analyze all knowledge needed to support the assessment of the conservation conditions, planning the restoration actions, and finally documenting the restoration results.
The data archival will be based on an accurate and high-resolution representation (2D, 3D, data of scientific investigation devices; usually multiple representations are endorsed and used). The system will allow Curators and Restorers to store and interlink all needed knowledge. This will entail:

- Basic national catalogue info
- Bibliography of literature related to the artwork and the story of previous restoration actions (text/reports and data)
- High quality digital representation of the artwork in its initial conditions (before restoration). This can be implemented with 3D models or even with 2D representation, when the artwork can be faithfully represented in the 2D domain (e.g. a painting or a fresco which do not have issues of layers detachments, which would require a 3D representation to be faithfully represented). Accurate digitization and reproduction of the colour and surface reflection characteristics are usually mandatory.
- Results of scientific diagnostic investigation analysis (multiple formats, possibly geo-referenced to the area of the object which was subject for the investigation)
- Results of inspection and study (annotations over the entire or selected parts of the artwork), to preserve all evidence found to support curator/restorers hypothesis and to document the study process. Documentation of the status entails the creation of a number of drawings characterizing the conservation status (which should be possible to draw directly on the digital clone, being it a 2D or 3D representation).
- High quality digital representation of the artwork in final conditions (after restoration) with 3D or 2D representation. Accurate digitization and reproduction of the colour and surface reflection characteristics are usually mandatory.

Here the focus is not just in documenting and archiving multiple heterogeneous documents, but also identifying/storing/visualizing connections among those assets; the digital clone of the artwork could be the spatial index for structuring and presenting all this knowledge. A cooperative approach will be enabled.

2.4.8 Why not a single digital clone system for curators/scholars/restorers?

The above three systems might have many functionalities in common, so an immediate query is why not a single system to support all three communities?
The answer is related to usability. It has been proved that complex and sophisticated systems are complex to master and to be endorsed by non-technological communities. A clear example are sophisticated CAD systems which have been used in many restoration actions [86] in the recent past, rising issues on the capability of Cultural Heritage experts to manage them.
The road to success is to provide a single community with exactly the set of functionalities needed to do a job, without the burden of a complex GUI which should provide access to tens or hundreds of unneeded features.
More the system is fitting well to the needs of the community, the easier the Cultural Heritage expert will master it and, at the end, will use it massively in daily work.
This was the experience granted in a complex restoration project where a documentation system was designed following exactly the needs and the set of functionalities requested by curators and restorers [87].

But this does not mean that these three systems should be designed in an independent manner. They should be designed in a modular manner, using common GUI and basic functionalities (e.g., annotation feature, visual analysis), thus facilitating a Cultural Heritage operator to move from one system to the other. Moreover, it should be possible to incorporate part of the data used in one system to initialize another one (e.g., an art historians willing to start a research on a specific artwork will initialize its digital clone from a proper subset of the data stored by the museum curator responsible of the same artwork; if this same artwork will have to undergo a restoration, the restoration clone should be initialized with the data already available in curator's or scholar's clones, if any).

2.4.9 Tools for designing / testing a new museum organization or a temporary exposition

Designing a new museum, or just a portion of an existing museum, or a temporary exhibition, requires a complex job encompassing several activities which are strictly related to the 3D space and related visitors' perception (subdivision of the exposition space, planning the visiting path, distribution of the artworks and of the didactic materials, lighting setup, etc). The preliminary design can be based on digital 3D technology and its preliminary evaluation can be done using interactive navigation (here the use of HMD can be ideal).
All this is nowadays possible on a technological side, but still very complex and high cost if implemented using commercial architectural CAD tools. But interior museum spaces are often quite simple from an architectural point of view, designing the exposition space often is just introducing walls and openings which allow to create a virtual path and smaller spaces inside a larger physical exposition space, displacing the artworks and setting a proper illumination.
Therefore, a tool for designing and providing rehearsal of a new exposition container can be implemented following the needs of this specific application context, focusing on just the required set of features, thus reducing usage complexity and lowering the bar of the skills needed to manage the system.

2.4.10 Tools for monitoring visitors' activity in museums, archaeological, monumental sites

Common queries of museum curators are how do the public visit the museum, how do people move around and which specific content they pay attention to [89]. The availability of an infrastructure for tracking each single visitor allows to get a huge quantity of data on effective museum usage and fruition, which can be used to evaluate the effectiveness of the current museum organization, to

understand which are the artworks of main interest (maybe also subdividing visitors in a number of categories), to reduce possible congestion of specific regions, etc.
Visitors' detection and tracking is nowadays possible using several different technical approaches [90]. But this is a quite complex technical task that museums often outsource to ICT companies.
The provision of a service, based on off-the-shelf technology, can be very useful for museums, since specific museum needs can be easily generalized and a common solution might be provided for data gathering and for supporting the analysis of the data.

2.4.11 Tools for producing interactive content or installations for museums

Museums often produce interactive content, designed to support visitors' increased comprehension of the artwork on stage (didactic purposes) or to entertain visitors. These are usually interactive multimedia presentations, often based on a story and implemented using a (linear) storytelling approach. The platform used to show the content can be a simple interactive presentation device (a PC, or a tablet, usually driven by a touch-based interface) or more sophisticated immersive devices (large screens, VR rooms, head-mounted displays, specific interaction languages and related devices). Adopting an interactive approach allows us to profit from the characteristics of sophisticated media (3D models, 360 images, etc.).
The most common approach so far is *single theme – single museum*: the interactive content is designed to tell the story of an historical context or an artwork of interest for a specific museum; then, the content is on stage only in a single museum. This approach increases substantially the cost impact, since everything is designed by scratch and presented only to a single museum audience. Most of the productions are planned in the framework of non-permanent exposition, since those events usually have a budget (often granted by a sponsor) and a specific focused theme. But designing for a temporary exhibition makes content production a transient job (to be implemented in a short time and alive for a short time).
How can we break this trend? There are some technical solutions which could alleviate some issues:

- Develop content with authoring tools enabled to present content on web-based platforms (offering the possibility to embed interactive visualization tools for visual media in standard web pages) , allowing to break the physical walls of the museum moving the content on the web, thus widening the potential public.
- MM content is built on top of basic components: the digital representations of works of art, architectural scenes, images (modern or historical photographs). The availability of a public archive of components will reduce the initial digitization cost of a MM production, since a considerable part of the required models/images will be already available and only a few specific artworks will have to be digitized.
- The design of MM content will be largely simplified by the existence of common authoring tools offered as services by the European Collaborative Cloud for Cultural Heritage, possibly easier to use than existing full-fledged MM content authoring systems (e.g., the Unity platform). Web-based content and related GUI should run on both desktop PCs and mobile devices, including the development of mobile apps.
- Those authoring tools should also offer the capability of customizing the presentations on more sophisticated presentation contexts (e.g., VR or AR systems).

2.4.12 Tools for managing bibliographies

Compiling (annotated) bibliographies is the first action in many Cultural Heritage professional activities, as well in many didactical uses. Professionals usually start from a specific artwork under investigation, or from a research question, and try to recollect all related knowledge in the form of an annotated bibliography. The work is often tedious and includes: finding and recollecting related works,

inserting bibliographic references in a list, evaluating each paper and inserting annotations in the bibliography.
This work is usually published as a component of the monographs/papers/reports produced in the framework of a research job or of a didactic exercise. But it has a value by itself and should be preserved and made accessible to the community.

Digital resources and tools could help professionals in several ways in this activity:

- providing a connection to bibliographic references available on the web (to retrieve citation references, digital copy of the papers);
- providing support in structuring and managing the annotated bibliography (authoring system supporting the creation, editing, and annotation of a bibliography, which should be sharable with a restricted group of co-workers or open to the large community);
- providing support in searching related papers (given a topic or a specific artwork), using current AI technologies working either on text or on images (automatic processing of the Bibliography of selected papers to retrieve related works; us of images in papers and text to evaluate the fit of a given paper with the subject of the research). Most of the tedious work can be therefore automatized, focusing scholar's contribution to the evaluation and validation of those results, augmenting the scholar's capability to analyse the huge number of references available on each subject.

2.4.13 Mixing media in visual presentation and navigation/analysis

A few pioneering experiences in the Cultural Heritage domain have already experimented the effectiveness of using *multiple media* in the visual presentation or analysis of artworks (e.g., interconnecting 3D and text to tell the story of an artwork, or intermixing 3D navigation and panoramic images to present both spatial and colour/reflection characteristics in a more accurate manner, or using 3D and images in the documentation of restoration) [88].
But this technology domain still requires to be extended with further research. To mention a few lines: designing new visualization methodologies based on mixed media, integrating interfaces and creating a common cross-media interaction language, and finally providing usable tools for the Cultural Heritage community.
One domain where this approach would be ideal is the digitization of *non-static heritage*. Modern heritage is often non-static, a clear example is the case of industrial heritage with many complex machines and processes that have to be cloned digitally. But producing a digital clone of a complex machinery can be an excessively expensive action. Conversely, we can use different media to represent in a cheaper and easier manner the dynamical essence of an asset: e.g., using 3D models to sample the static characteristics of a machinery and adding video and sound to enhance the perception of the dynamic experience.

2.4.13 Supporting archival and documentation of archaeological excavations

Archaeological investigation is an activity that produces a relevant amount of digital data, survey documentation, and documents presenting/narrating/interpreting the findings.
When carrying out an archaeological excavation, almost all countries enforce legal obligations regarding the archival of the produced data (photographic data, 3D models, drawings, plans, GIS data), and the creation of periodic reports (that might be yearly, after every campaign, or even after each month/week of work, depending on the circumstances).
These requirements have an impact on both the academic domain and the commercial archaeological companies; the latter have a considerable share of the overall activity, considering that the vast majority of excavation are actually instances of Rescue Archaeology (https://en.wikipedia.org/wiki/Rescue_archaeology). In spite of this, usually the academic domain

generates and stores more data, because of the additional needs for a detailed study and the use in teaching.

Therefore, it would be important to support both domains with an archival and documentation system able to:

a) *archive* the excavation data in a structured way, and
b) provide *authoring instruments* to access the stored data and to reuse them to build reports, interpretations and teaching material.

The *archival* side should support a diverse array of data (2D, 3D, geographical data, audio & video), but, more importantly, it should be designed exploiting all the advantages of the semantic web. Using established ontologies and vocabularies to define the data structures and relationships between the data would grant a new level of interoperability. A well-defined, semantic-enriched archive would provide an unambiguous and easily interpretable data corpus, able to withstand the test of time; it would be possible to integrate and connect archives from different institutions and countries; and the semantic information layer would make possible for AI technologies to access and "learn" from the archive. This point, in particular, would open up interesting scenarios, as it would solve the main issue in the use of Deep Learning and AI on Cultural Heritage data that is the lack of annotated dataset for training.
An important feature would also be the possibility to provide citation capabilities for the archived data, to facilitate publication, sharing and data reuse in such a way to preserve the integrity of the sources (and, of course, its property).

The *authoring* part should provide a way for the user to build "narratives" starting from the archival data. This activity is an essential part of the process of knowledge extraction in this field. By "narratives", we mean all those documents that use, as their building blocks, the data entities contained in the archive: the mandatory periodic reports, the formulation of new interpretations or the confutation/discussion of the already existing ones, the preparation of scientific papers and publications, the set-up of teaching/didactic material, schemes and graphs. The users should be able to semantically access and link in these documents the archival contents, possibly using dedicated, visual authoring tools. As a side effect of the semantic nature of the system, also all the produced content should be semantically annotated and made available as new layers of data.

Looking closely, this is basically what a data-science platform should provide. However, the needs and peculiarities of this field would require a careful design and implementation of the features sketched above.

2.4.14 Considerations on the development of tools

Tools matrix

1 Least expensive Big demand	2 Most expensive Big demand
3 Least expensive Poor demand	4 Most expensive Poor demand

Considering ideas for tools, the project should focus first on those that meet biggest demand and require least budget to develop, second on those that meet biggest demand and are more expensive, third on those that serve a smaller use base but are cheaper to develop. Ideas for expensive tools that serve a small use base should be dismissed.

The approach could be:

- Start with a quite complex project (with a big budget) aimed at building all the basic components of the infrastructure plus one or two applications;
- Follow up with smaller projects (each one with a small budget) each of them aiming to implement just one tool on top of the Cloud.

Perhaps **finding and presenting data** can be a native application in the first project. Focus at first on a complex platform capable of securing and managing large amounts of (research) data. A second application could be the "**Digital Clones for Museum Curators**" presented above, again in the first project.
The design of the Cloud should be sufficiently smart to accommodate future use cases (follow-up applications). This vision is the base for the proposed architecture and implementation (see section 3.3 for a more detailed description).

Many programs for backing up large amounts of data no longer receive updates to run on local server-systems. Users are thus often forced to switch to a commercial cloud. As a result, the security of the data is no longer guaranteed, because those cloud are often operated under American jurisdiction and commercial aspects.

Part 3 A EUROPEAN COLLABORATIVE CLOUD FOR CULTURAL HERITAGE – AN IMPLEMENTATION PLAN

3.1 European Collaborative Cloud for Cultural Heritage architecture

3.1.1 Main motivation for designing a European Collaborative Cloud for Cultural Heritage

The Cultural Heritage domain lacks a European Collaborative Cloud for storing, accessing, using and documenting digital twins, for supporting the activities of the digital continuum and for connecting all actors in the digital ecosystem. As already discussed in section 2.2, existing clouds provide different services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS), see [1]. In IaaS, Cloud Service Providers like DropBox and others offer the infrastructure and network components as a service while users can implement and use their software applications using the provided virtual resources. In PaaS systems like Windows Azure, the Red Hat OpenShift or the Aptana Cloud, users can develop their own applications using cloud-related development environments. On the other hand, SaaS cloud systems like Zoho and Google Docs offer closed applications that are hosted, maintained and managed by cloud service providers through a central location. However, existing SaaS systems are too closed to be suitable for the interests of stakeholders in the Cultural Heritage domain, while IaaS and PaaS systems lack integration facilities, dynamic adaptation to user needs, and long-term sustainability among other shortcomings.

Designing a European Collaborative Cloud for Cultural Heritage will be crucial for the present and future needs of all involved stakeholders. Moreover, having this Cloud will foster initiatives from all corners of Europe, connecting experiences, sharing activities, promoting a European perspective of Cultural Heritage, and cultivating the need of "going digital" in many Cultural Heritage institutions across Europe that have not yet entered the digital world. The European Collaborative Cloud for Cultural Heritage is not only necessary for the community, but is also essential for its expansion and unification at European level.

As already discussed in Part 2, proprietary clouds no not guarantee the long-term EU objectives on Cultural Heritage, because changes and updates would be decided by corporations that are out of the EU control (data servers of proprietary clouds would be probably located out of the European area, they are not focusing on the Cultural Heritage domain due to the current small business potential), their extensibility in the future being also unclear.

Instead, the European Collaborative Cloud for Cultural Heritage design should be user-driven and adapted to the needs of the community. It should be usable in broadband **networks** but also for museums away from very high-speed communication streams; especially those in rural areas. The Cloud should be intrinsically modular, dynamic, evolutionary, and extendible. By being based in Europe, it will also facilitate the joint analysis and presentation of multiple and rich duplets made up of artworks together with their digital twin. It may also promote and support the creation of stories and transnational experiences connecting the now dispersed elements of European Heritage.

These requirements lead to an **open-source framework**. The European Collaborative Cloud for Cultural Heritage design should derive from the community needs and from the analysis of present open-source alternatives and open-source cloud-based collaboration platforms, like NextCloud and ownCloud, see [96] and [98] (an extensive comparison between NextCloud and the most popular closed-source services is available at [97], giving relevant information on the drawbacks of existing cloud systems).

The European Collaborative Cloud for Cultural Heritage must provide a data space, offering also computing resources for running user codes and a virtual research environment to support cooperative work of groups of scientists and professionals. It must provide a repository with storage and retrieval services including portals indexing all existing tools and digital models, indexing catalogues, access rights, provenance and other relevant information. Tools to interact with the data must also be provided, including interactive visualization tools and advanced data analysis. The European Collaborative Cloud for Cultural Heritage also intends to facilitate the daily work of cultural heritage professionals, which will be one of the conditions for its success. Finally, the European Collaborative Cloud for Cultural Heritage design must guarantee assessment tools, software tools to implement the Cloud evolution, and long-term access including persistent data preservation.

3.1.2 Overall European Collaborative Cloud for Cultural Heritage architecture

Cloud requirements

According to the conclusions in sections 2.1 and 2.2, we can list a number of requirements that should drive the design of the European Collaborative Cloud for Cultural Heritage:

- The European Collaborative Cloud for Cultural Heritage should provide **web-based software** and **services**
  - The Cloud should equally address professionals and researchers
  - The Cloud design should be user-driven, based on the needs of the community
  - The Cloud should be open-source, being distributed and inter-operated, and having a basic platform of medium size. It should include appropriate network connectivity and management tools. Having an open-source architecture, it is a guarantee to avoid misuse of the data with private interest goals.
  - As mentioned in section 3.2, the European Collaborative Cloud for Cultural Heritage should provide ontology and vocabulary services with APIs for re-using them, data services for publishing data models and datasets for their re-use and application development, and software tools for aggregating, producing, publishing, and analyzing datasets. The Cultural Heritage should include best practices for cataloguing, publishing and using Cultural Heritage data.
  - The European Collaborative Cloud for Cultural Heritage should support user authentication (single user/groups of users) and private and public data management. The authentication of users and user groups must allow the definition of who can use each of the digital models and services offered by the European Collaborative Cloud for Cultural Heritage, also ensuring the traceability of all actions carried out on it.
  - The Cloud should be intrinsically modular, dynamic, evolutionary, and extendible.
  - The Cloud should be adapted to broadband networks but also to museums and users with low-speed communication streams; especially those in rural areas.

- The European Collaborative Cloud for Cultural Heritage must offer a **repository** for the online archive of **digital twins**.
  - Digital twins will be enriched digital models including data, metadata, paradata, and provenance information.
  - The Cloud design should be distributed, also using national repositories and digital networks. It should include a transparent interface with existing local (national) Cultural Heritage data storage systems. Data coherence should be guaranteed. The system should allow national communities or institutions to install and configure their

own local Cloud (if needed for internal policy reasons). Users' access and queries should be resolved in a transparent manner on the entire distributed Cloud.
    - The Cloud should include **security facilities** like data encryption, directory services integration, authorization levels and auditing events for specific resources, and third-party cloud evaluations (2).
    - The Cloud design should foster data fidelity by promoting the storage of measurable digital models that encode local fidelity information along with the geometric and appearance information, as detailed in section 3.1.

- The European Collaborative Cloud for Cultural Heritage should provide **portals** for **data retrieval** with a user-friendly and unified graphical user interface (GUI). Portals should:
    - Offer visualization, interaction and data analysis tools (see Part 2).
    - Deliver a unified management of public and private data.
    - Be open to small organizations and museums, offering them user-friendly tools and services.
    - Connect activities from different Cultural Heritage actors, fostering a digital continuum of projects and interests.
    - Bridge actors, activities, objects and twins, to create a European Cultural Heritage ecosystem.

- The European Collaborative Cloud for Cultural Heritage should ensure **long-term use** through a reliable and stable **assessment** mechanism:
    - The Cloud should support an evolutionary model: the future European Collaborative Cloud for Cultural Heritage evolution should be based on measurable **success indicators** (qualifying both the entire Cloud and each specific component), and should provide a continuous monitoring of the use of data and tools by the community of stakeholders. Successful components should be maintained and extended, the parts which are not backed by an active community and under-perform should be discontinued.
    - It should be possible to scale the platform according to user needs and real usage indicators; this includes the bandwidth available, the number of foreseen users, and a forecast of the evolution of the number and size of the digital assets.
    - Cloud assessment should be independent and long-time stable. It should be based on the evolving community needs, also guaranteeing European public interests in Cultural Heritage. The governance model of the Cloud should warrant the independence of this essential process, also guaranteeing that all required changes are carried out.

Proposed architecture: basic principles and structure

The architecture of the European Collaborative Cloud for Cultural Heritage should be designed with a long-term vision according to the needs of the Cultural Heritage community, being evolutionary, extendable, and dynamic. The design of the Cloud architecture should be based on some basic principles:

- The European Collaborative Cloud for Cultural Heritage should ensure compliance with all requirements already presented above.
- The European Collaborative Cloud for Cultural Heritage should provide a data space, offering also computing resources for running user codes and a virtual research environment to support cooperative work of professionals and researchers.
- The Cloud should provide a repository with storage and retrieval services including indexing all existing tools and digital models, catalogues, provenance and other relevant information.

- Tools to interact with the data must also be provided, including interactive visualization tools and advanced data analysis.
- The European Collaborative Cloud for Cultural Heritage should adapt to evolving data types and new data models by including an intermediate interface based on data dictionaries to disconnect multiple and evolving data implementations from tools and applications, and by offering any standard and stable API to these tools.
- The European Collaborative Cloud for Cultural Heritage should also offer an API for extending it with novel tools and applications like training materials, documentation, support for living labs, etc. Easy plug-in of new tools, built on top of the European Collaborative Cloud for Cultural Heritage data layers, is essential. Software tools for the European Collaborative Cloud for Cultural Heritage evolution should be provided.
- The European Collaborative Cloud for Cultural Heritage should include (see subsection 3.2.1) libraries providing: authentication of users; common and uniform access to the cloud data; visualization libraries; common GUI components; and libraries for monitoring and keeping memory of the use of data and of the specific applications. The authentication of users and user groups must define who can use each of the digital models and services offered by the European Collaborative Cloud for Cultural Heritage, also ensuring the traceability of all actions carried out on it.
- The European Collaborative Cloud for Cultural Heritage should enable the implementation of private areas that, upon a decision of the owners (either before or once the related specific project is terminated), can become public (examples are restoration actions or archaeological campaigns, where the initial phase is private and restricted, while the final results and part of the discovery process should later become public).
- The European Collaborative Cloud for Cultural Heritage should include instruments for evaluating the success of tools/services (which should be the basic instruments for implementing the assessment system and the evolutionary approach). It would be wise to foster users' fidelity, to ensure increasing Cloud use and that the data available will progressively and substantially grow in size along all the life of the Cloud (see subsection 3.2.3).

As shown in the next diagram (Figure 2), the European Collaborative Cloud for Cultural Heritage should be structured in different layers, to successfully address the listed requirements.

From bottom to top, we first have the digital data, models, metadata and paradata. The second layer, named "Data Interface", shows a data API providing a unique interface to the tools and application on the top of it, also supporting data in different and evolving formats. This second layer ensures that all tools and applications will always support all data types in the Cloud, both the old ones and the future ones, with a unified API access. The third level from the bottom includes the basic infrastructure and libraries to support creation and plugin of apps (including visualization applications), support for user authentication, and support for the implementation of long-time assessment tools, which requires gathering data from all raw data and all Apps, also from the external ones. Finally, the top level includes all tools and applications (digital twins, digital continuum, digital ecosystem). The first step of the implementation (basic Cloud consortium) would design and implement a first version of the three bottom layers plus some few applications to test the initial Cloud system.

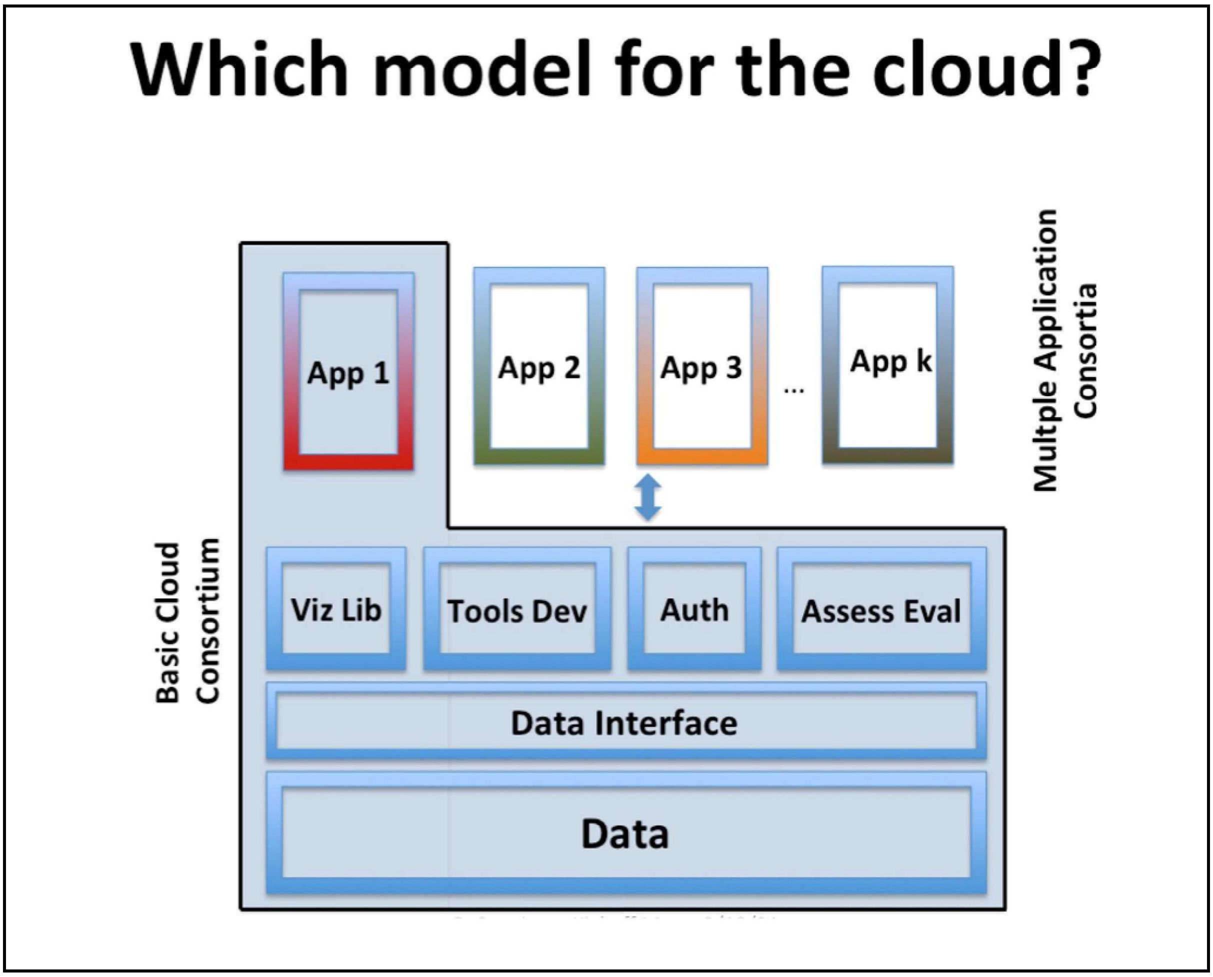


Figure 2: The diagram shows the architecture of the European Collaborative Cloud for Cultural Heritage, organized in several layers.

Data model and Data Components Needed

The data model should be designed to support and to enable a number of characteristics:

- Specific data representation schemes or data types used in the Cloud should be open or based on international standards.
- Each **digital asset** (digital twin representing a work of art) should be paired by **metadata** (describing the artwork represented, the conservation institution...) and **paradata** (documenting who produced the model, when, the technical instruments used in digitization, the acquisition parameters, etc);
- **Derived** digital assets coming from all cloud-related activities (digital continuum) can be produced from each digital asset; they should also be stored in the Cloud, keeping reference to original metadata and paradata (**provenance**), plus adding specifications on how they have been derived from the original asset (to allow to evaluate data **integrity** wrt. the basic original asset).

- The metadata schema and data types should enable storing information on **data quality** and **provenance**, allowing to trace the process from digitized originals to derived assets (see subsection 3.1.3)
- Moreover, since the European Collaborative Cloud for Cultural Heritage also aims at increasing data reuse, the data model should also allow storing data on the **different uses** of a specific digital asset. This info will be updated dynamically, following the life of the digital assets, being partly based on the assessment instruments that will generate information on the use (see section 1.4) and being stored as metadata. Therefore, it should be possible to query the Cloud which are the uses of a specific digital asset or of all its derivations.
- Any layer of the European Collaborative Cloud for Cultural Heritage should be able to evolve to support future needs of all users. When modifying digital data at the bottom layer, coherence of the overall Cloud should be always preserved.

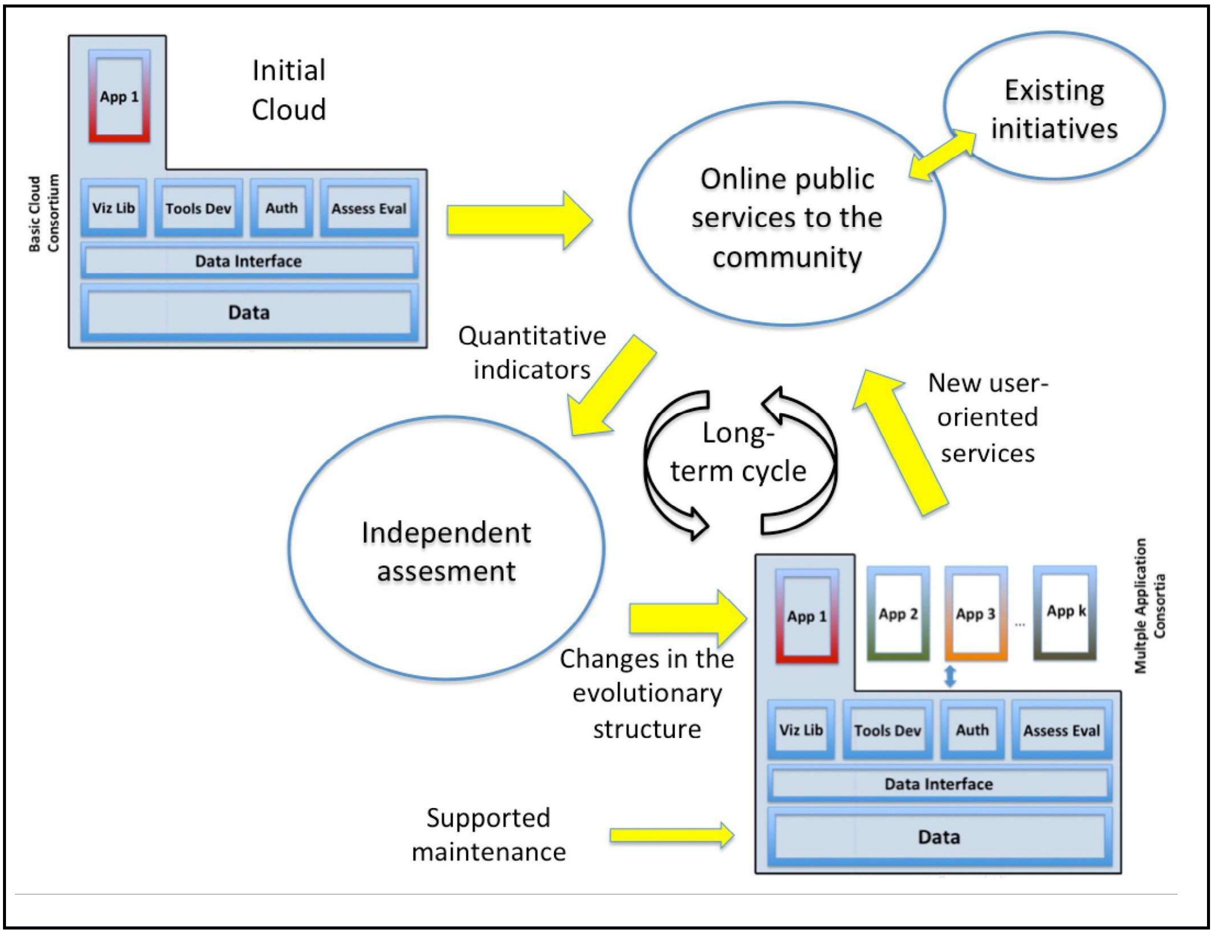


Figure 3: The evolving nature of the European Collaborative Cloud for Cultural Heritage is presented here.

Assessment and testing

The test and assessment of the European Collaborative Cloud for Cultural Heritage has to be performed both during the Cloud development and at the final stage of the project (iterative testing).

Moreover, assessment should be repeated for all the lifespan of the European Collaborative Cloud for Cultural Heritage (e.g., every 24 months). It should consider both the basic infrastructure and the native applications built by the original European Collaborative Cloud for Cultural Heritage project and the other applications built on top of the European Collaborative Cloud for Cultural Heritage by other subsequent ancillary EC projects.

The features and tools should be tested including both technical tests (evaluating correct functioning of the features implemented) and user-driven evaluations (to assess that the European Collaborative Cloud for Cultural Heritage fulfills the users' expectations and to check its practical usability with the prescribed user communities).

Independent assessment based on an objective analysis and metrics of the use of the European Collaborative Cloud for Cultural Heritage data and tools is essential, being the driving force of the future Cloud adaptation and updates. This is shown in Figure 3.

The initial European Collaborative Cloud for Cultural Heritage (top-left in Figure 3) will provide a number of basic services to the community, including standardized access to the data, basic libraries and some initial tools and applications as already discussed. Then, long-term maintenance and adaptive improvement of the Cloud must be user-oriented through an independent assessment body. The triangle formed by the user community, the independent assessment, and the future application consortia that will continue adapting and upgrading the European Collaborative Cloud for Cultural Heritage becomes a guarantee of the future survival and usability of the overall system. This triangle is user-driven, with the independent assessment system being always receptive to the evolving needs of the professional stakeholders. Cloud usage should be evaluated in an independent way through objective indicators and living labs, leading to the specification of new changes to be introduced in the Cloud structure. These changes will be implemented by assessment- driven application consortia that will hopefully improve the online public services that the European Collaborative Cloud for Cultural Heritage will offer to the community. This triangle including users, independent assessment and consortia updating the Cloud should be a guarantee of a long-time benefit for the community. Of course, the European Collaborative Cloud for Cultural Heritage will be linked to existing European initiatives, as also shown in the diagram, including a suitable business model (see section 3.4) to guarantee the essential assessment and maintenance activities in the loop, and integrating both national networks and small organisations and museums to help them evolve to the digital universe. Last but not least, the European Collaborative Cloud for Cultural Heritage could (and should) become the new European instrument to connect Cultural Heritage at the continental level, making resources integrated and interoperable.

Living labs can be organized with related user communities, to contribute both to dissemination goals and assessment.
Living labs can also be instrumental in creating a strong interconnection between the European Collaborative Cloud for Cultural Heritage consortium and all the other consortia responsible for the other projects designing add-ons for the European Collaborative Cloud for Cultural Heritage (other tools and extensions). Living labs can be the context where integration pilots are tested and (users) experience is exchanged.

3.1.3 Assessment of the potential risks

Risk: **Not enough support from communities, data uploading is below the expected figures.**

Countermeasures: Provide pilot projects for running the first pioneering activities, to show to stakeholders how to use the European Collaborative Cloud for Cultural Heritage with practical examples. Plan an intense training and dissemination program. Involve key stakeholders in Living Labs and training activity.

Risk: **Failure in providing consistent and easy-to-use GUIs among the different cloud tools and resources**

Countermeasures: Usability is key to success. Design of tools and of related GUI should be user-driven, by including groups of qualified users in the design phase and in the assessment. Interaction and collaboration among the European Collaborative Cloud for Cultural Heritage consortium and the different consortia active on the European Collaborative Cloud for Cultural Heritage (thus including subsequent projects) is another key factor. This should be explicitly considered at management level (a management body should be created by the European Collaborative Cloud for Cultural Heritage consortium, including all partners and selected stakeholder, and open to contributions and members representative of other EC projects contributing to the Cloud). Common policies and guidelines for the design & implementation of GUI should be drafted by the European Collaborative Cloud for Cultural Heritage consortium.

Risk: **Failure in providing an independent, long-term assessment and evaluation of the Cloud**

Countermeasures: As independent assessment is an essential element of the European Collaborative Cloud for Cultural Heritage, ensuring its quality is imperative. The internal mechanisms of the assessment body should be carefully designed, ensuring that public interests of the community and European interests are well protected. Perform periodic surveys to the broad community of Cultural Heritage stakeholders.

3.2 Support for implementation of tools/services using the European Collaborative Cloud for Cultural Heritage data and resources

We describe here the architectural components that shall be provided by the European Collaborative Cloud for Cultural Heritage to support:

- easy plug-in of new tools, built on top of the data archive and low-/mid-level functionalities of the Cloud;
- monitoring the real use of the resources provided by the Cloud (both numeric figures and qualitative assessment);
- enforcing users' fidelity, to ensure increasing Cloud use and that the data available will progressively and substantially grow in size along all the life of the European Collaborative Cloud for Cultural Heritage.

3.2.1 API for extending the European Collaborative Cloud for Cultural Heritage with other applications/tools

The European Collaborative Cloud for Cultural Heritage should host applications, which are (interactive) tools working with (preferentially) the data stored on the Cloud.
So far, we have used both the terms tool or application. But these can be implemented in two different ways:

- web-based applications, with the GUI contained in a standard web page (for example, this will be the format of the application providing ingestion, search and retrieval services, see section 3.3);
- stand-alone applications, requiring heavy computing resources, intense data throughput, connection to acquisition devices, etc. (for example, this will probably be the case of

applications implementing new digitization approaches, see section 3.3). These resources can be offered as downloadable SW (the user will run it on his computer) or as SAAS resources (see section 1.2 for the definition of SAAS).

What differentiate this case from the usual cloud platforms offering computing services (thus, able to run user codes or applications) is that the European Collaborative Cloud for Cultural Heritage applications will:

- access data stored on the Cloud repository (those will be the preferred input for the apps);
- use a set of European Collaborative Cloud for Cultural Heritage libraries providing: authentication of users; common and uniform access to the Cloud data; common visualization libraries; common GUI components; and libraries for monitoring and keeping memory of the use of data and of the specific app (data to be ingested in the Cloud);
- upload (when needed) the results produced on the European Collaborative Cloud for Cultural Heritage; the latter can be new data or enrichment of the data already stored there.

Therefore, support material should be provided for applications developers (most of them external to the consortium responsible of the European Collaborative Cloud for Cultural Heritage design and implementation) to make them fully aware of the data model used to encode and describe the different types of data stored in the European Collaborative Cloud for Cultural Heritage, and the libraries provided. This implies that an important and critical effort should be dedicated to the production of training materials (system documentation, how-to-guides, courses, living labs).

The initial applications provided by the European Collaborative Cloud for Cultural Heritage will aim to a double scope: 1) provide some basic capabilities for the Cloud, going beyond pure data encoding and storage features; and 2) being a living example for external application developers (for example, for the consortia winning subsequent EC calls aimed at the completion of the European Collaborative Cloud for Cultural Heritage features), providing them some practical examples and demonstrating how an application should be built.

3.2.2 Support for monitoring data and application reuse and impact (all-life assessment)

Data access: for each digital asset, the platform should maintain counters of how many times this specific asset has been "accessed". For example, for visual data assets access means how many times the file has been opened and visualized with the platform visualization tool (or library component). This should become one of the metadata associated with the visual assets and should be subject to queries (to be decided if this will be open to everybody or restricted to the Cloud administrators). The platform should also provide instruments for producing statistics and summaries on data usage.
Data use is not just "accessing" a specific data item, but also using it in other projects. This will be more complex to store and resume for assessment purposes. It can be a numeric figure, but should also be qualitative (i.e., being related to the "quality" and "success" of the context where the data is reused). For each reuse action we will need to ingest in the European Collaborative Cloud for Cultural Heritage also some data to characterize the specific context (textual info, a link to a related web page, etc.). The concepts of "research project", "museum exposition", "restoration action", "museum installation", etc. are all concepts which should be part of the items represented/archived in the European Collaborative Cloud for Cultural Heritage (since our community needs to describe and preserve data for all these actions). Therefore, we should create links between any (reused) digital object and all the actions where it has been used.

Tools use: here the simplest indicator is frequency of use. This can be measured as how many times the specific tool has been activated by users (total number of activations, no. of users, average no. of activations/user, etc.). Another indicator is the type(s) (qualitative indicator) and number (numeric indicator) of actions, to which it has been adopted.

Tools vitality: each single tool could be either frozen (with a very minor activity on its evolution, maintenance or extension) or actively maintained and updated (with frequent commits of the related code, implementing bug fixing or extensions). This indicator can be easily reported by the Cloud platform (it is common on SW development platforms). It is not an indicator of quality, but of the level of support it receives by the development community.

3.2.3 Support for enforcing users' fidelity

The consortium in charge of the European Collaborative Cloud for Cultural Heritage design and implementation should also consider activating measures to enforce users' fidelity. The success of the European Collaborative Cloud for Cultural Heritage will strongly depend on the motivation and contributions of a large number of stakeholders/users. The actual data population of the Cloud is in their hands, it will depend on the contributions of external users.
Any effort in making visible the contribution of users and rewarding them will be beneficial for the success and visibility of the Cloud.
The European Collaborative Cloud for Cultural Heritage Consortium should be very creative on this subject, devising instruments for creating and enforcing the community of users.
This subject is mentioned in this subsection, because some of these policies may need some technical support such as:

- Establishing badges to recognize categories of premium contributors (silver, gold, platinum), according to the volume of contribution to the European Collaborative Cloud for Cultural Heritage. To gather the related figures, we need to be able to detect the users (either single persons or institutions), which go beyond specific targets (no. of digital assets contributed to the Cloud; frequency of use of the Cloud tools, etc.);
- Awards presented to institutions/users who have produced works or results of excellent value (some sort of yearly awards). These awards require peer evaluation, but could also take into account figures concerning the number of accesses to the pages or resources of those specific results or any explicit "like-style" assessment produced by the community (using some related European Collaborative Cloud for Cultural Heritage enabling feature).

3.3 Possible tools to be built on top the European Collaborative Cloud for Cultural Heritage

This section presents a list of potential tools, to be developed on top of the European Collaborative Cloud for Cultural Heritage.

Following the vision presented in the previous sections (in Parts 2 and 3), the overall functionalities that a European Collaborative Cloud for Cultural Heritage should fulfill are quite complex, differentiated and sophisticated. Therefore, we think that it would be too complicated to endorse a policy, where the entire set of tools or applications should be designed by a single consortium in the framework of a single initiative. As it has been already proposed in this report, we endorse a progressive and evolutionary model, in which the European Collaborative Cloud for Cultural Heritage should be constructed in several phases and with the (well-coordinated and integrated) contributions of several consortia. In consequence, first, the basic tools which should be provided by the initial European Collaborative Cloud for Cultural Heritage consortium are listed, and then the other additional efforts, by organizing them on a possible time scale (somehow related to the future multiple

EC Calls which will select the consortia and will fund those added efforts). We endorse here a three-year time frame, corresponding to potential calls to be issued by EC in 2023, 2024 and 2025 (and thus covering the overall time span 2023-2028).

The content of this section is consequent with the content of section 2.4 *Potential uses for museums or Cultural Heritage institutions*. Therefore, we do not repeat the content of section 2.4 here, but just recall in a short synthesis the purpose of the applications/tools, which were already described there.

3.3.1 Integrating (existing and operational) basic tools and services

The scope of the European Collaborative Cloud for Cultural Heritage is to provide common resources to Cultural Heritage communities at large.

Therefore, if the approach is based on *sharing principles* (data and tools), a first action of the European Collaborative Cloud for Cultural Heritage should be to make an inventory of the existing efforts and to select the ones that can be incorporated in the Cloud - since there is no need to *reinvent the wheel*. This could entail low- or mid-level libraries, and already operational applications or tools. In both cases some re-design will be needed, to allow third-party resources to be compliant with the data model and the specifications of the European Collaborative Cloud for Cultural Heritage. In the proposal preparation phase of a bid for the European Collaborative Cloud for Cultural Heritage, these third-party components should be preliminary individuated and resources should be allocated for their incorporation.

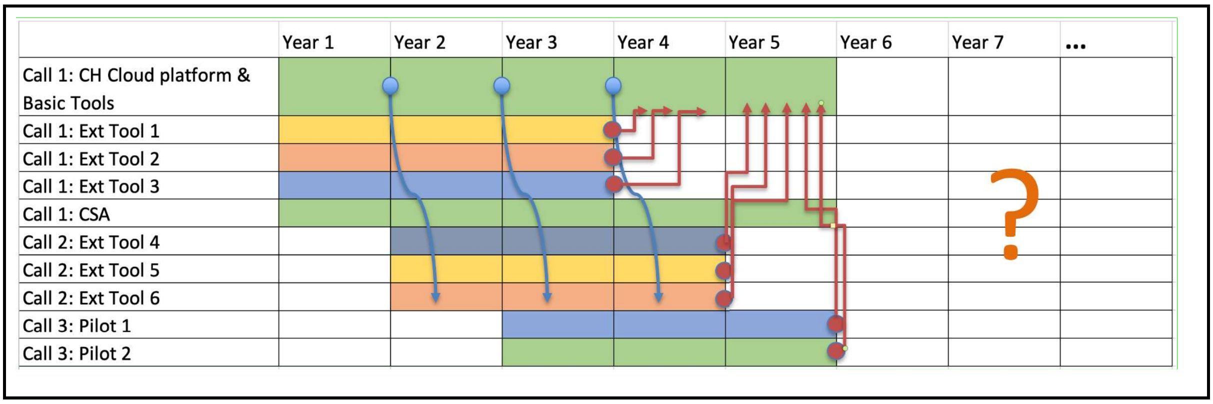


Figure 4: This scheme presents graphically the timings and the relation between the different projects contributing to the European Collaborative Cloud for Cultural Heritage venture. The basic building block will be the call for the basic European Collaborative Cloud for Cultural Heritage (including the overall infrastructure and two basic tools). This initial project should be paired with several other projects, starting in year 1, 2 or 3. Key milestones of the European Collaborative Cloud for Cultural Heritage project will play a critical role also for these other projects (see the blue lines in the drawing). These latter projects will contribute their results to the European Collaborative Cloud for Cultural Heritage in order to become applications integrated and offered by the European Collaborative Cloud for Cultural Heritage (see the red lines). A CSA action is planned, having the critical role of consolidating a community wider than the project consortia, ensuring a smooth collaboration and a user-driven design, implementing assessment policies (evolutionary model) and contributing to establish a sustainability policy. The question mark at the right side is to underline that a policy should

be defined and put in place to ensure the European Collaborative Cloud for Cultural Heritage would persist in time, remain operational and able to further evolve.

3.3.2 Native cloud-based tools and services (Time 1)

To reduce the extent of the work planned for the initial European Collaborative Cloud for Cultural Heritage consortium, we envision only two tools which should be designed and implemented in the framework of the first specific Call (with a project duration of 5 years and a maintenance plan with a following term), to be constituent parts of the initial European Collaborative Cloud for Cultural Heritage design (see Figure 4).
These initial tools should be selected to fulfill the general needs of our main community (museums).
Moreover, these tools will be provided by the European Collaborative Cloud for Cultural Heritage consortium as documented examples to explain to other consortia or external programmers how to develop sophisticated applications on top of the European Collaborative Cloud for Cultural Heritage low- and mid-level libraries and fulfilling the cloud policies (use of data, design or GUI, usability, documentation, etc.). Therefore, the scope of these initial applications will be both operative and didactic.

BasicTool 1: Data Interface (ingestion, search & retrieval)
This is the first tool which should appear on the time scale, since it is the basis for contributing material to the repository, for adding content, for searching and discovering useful digital assets which could be analyzed and reused.

BasicTool 2: Digital Clones for Museum Curators
Since the museum community is our main stakeholder, the tool dedicated to the management of collections (to structure, encode, store and analyze all knowledge needed to support curation activities) is our guess for the second native tool.

Moreover, another important task of the initial should also be to define a governance (open to the contribution of external stakeholder) and strategies/policies for the sustainability of the European Collaborative Cloud for Cultural Heritage after the end of the project (including establishing tight relations and synergies with existing infrastructures and other initiatives).

3.3.3 Additional tools and services (Time 1)

These actions might start in synchrony with the European Collaborative Cloud for Cultural Heritage call (same year), but will be implemented by other consortia. Therefore, we should select actions requiring a very loose integration with the European Collaborative Cloud for Cultural Heritage data model and services. The consortia in charge of these other actions should work in tight collaboration with the European Collaborative Cloud for Cultural Heritage consortium, to ensure that the decisions taken by the different consortia do contribute to the same vision and results could be easily integrated to the European Collaborative Cloud for Cultural Heritage.

Among the tools or services described in section 2.4, potential candidates to be included in the Time1 pool are:

ExtTool 1: Tool for managing bibliographies

This action will follow the current Open Science trend, will support a common, shared management of bibliographic references and related instruments; it will also add innovative tools for the semi-automatic construction of annotated bibliographies.

ExtTool 2: Tool for monitoring visitors' activity in museums or archaeological/monumental sites
This tool will provide instruments to monitor and evaluate how visitors perform their visit (how they navigate the museum space, timings, objects of interest and reactions; all these functionalities based on technologies able to track visitors).

ExtTool 3: Advanced digitization instruments
Instruments (mostly software) required for extending the digitization capabilities. Some possible focuses are: enabling more massive digitization, also devising new AI-based solution for sampled data integration and processing; solution for digitally cloning complex artworks (complex assembly, dynamic objects); improved representation of complex materials or to go beyond the visible spectrum; solutions for integrating and enriching digital twins with time-related information from sensor networks, which sample physical properties/states.

ExtTool 4: Tool for designing and testing a new museum organization or a temporary exposition
This tool (probably, a stand-alone application rather than a web service) will provide instruments to design a new exposition (defining the museums spaces in 3D and allocating the artworks in those spaces) and to test the related design (virtual navigation and rehearsal). It will have a loose interaction with the European Collaborative Cloud for Cultural Heritage (just the retrieval of the 3D or 2D digital clones of the artworks, to be allocated in the virtual exposition spaces).

CSA Action
Setup of an advisory and coordinating board who should interact with all consortia created for the overall European Collaborative Cloud for Cultural Heritage initiative, with the main tasks of:

- increasing the level of interconnection of the technical consortia with the Cultural Heritage community;
- contributing to the assessment (focusing more on the integration and interoperability aspects, checking congruence of GUI among different tools, verifying usability and practical effectiveness);
- monitoring the expectations of the community and supervising the user-based evaluation of the European Collaborative Cloud for Cultural Heritage resources;
- contributing to the future development agenda of the European Collaborative Cloud for Cultural Heritage. The CSA should also govern the interfaces between projects and existing Infra - DARIAH, ERIHS, EOSC, Europeana, etc.

3.3.4 Additional tools and services (Time 2)

Time 2 means that the actions planned in this group will start one year after the start of the European Collaborative Cloud for Cultural Heritage work. This entails that these actions should be selected as the ones requiring tighter integration with the European Collaborative Cloud for Cultural Heritage. Starting one year later will allow to consolidate to some extent the work in the European Collaborative Cloud for Cultural Heritage project and thus to have some of the European Collaborative Cloud for Cultural Heritage technologies in operation when those tier 2 projects will require them.
Among the tools or services described in section 2.4, potential candidates to be included in the Time 2 pool are:

ExtTool 5: Tool for metadata creation and enrichment

Complement human-driven data curation with approaches (ontology-based, linked data, AI-assisted, cross-languages vocabularies, etc.), which could speedup, improve and enforce the metadata creation and enrichment phases. Those semi-automatic features will be subject to human experts' supervision and validation. They will be key resources in enriching and enlarging scope, usability and impact of the European Collaborative Cloud for Cultural Heritage data.

ExtTool 6: Digital Twins for Art Historians/Scholars
This tool will allow Art Historians/Scholars to structure, encode, store and analyze all knowledge needed to support the study of the artwork. The digital clone will be the base for a number of activities (characterize, annotate, connect, compare, measure), which could be the base of cooperative and remote activities.

ExtTool 7: Digital Twins for restoration
The tool will allow to support the many actions needed in Cultural Heritage restoration projects, using the digital clone as a basic support for many cooperative actions (search and organization of previous knowledge, scientific investigations, data integration, data analysis, characterization of conservation conditions, annotations, document the process and the results).

ExtTool 8: Tools for producing interactive content or installations for museums
Design of simple tools, to enable the production of interactive content or installations for museums fruition. The focus is not in replicating full-fledged commercial authoring systems for multimedia production or computer animation, but to offer an easy-to-use and minimal platform for producing and sharing content among Cultural Heritage institutions. This should be aimed at the web communication channel and being delivered inside the museum (kiosks or mobile devices).

ExtTool 9: Enhanced interaction modalities
This project will study enhanced interaction modalities, to be contributed to the European Collaborative Cloud for Cultural Heritage by extending the mid- and high-level libraries for data visualization and analysis. The work could include: new approaches for mixing and integrating different types of media in visual presentation/navigation/analysis; new interaction modalities associated with innovative VR/AR technologies.

ExtTool 10: Technologies to enable enhanced IPR management
New technologies enabling an improved management of IPR and valorization of Cultural Heritage and its digital counterpart. This will include technologies to protect digital assets (e.g., watermarking) and to monitor their use (e.g., blockchain).

3.3.5 Additional tools and services (Time 3)

Time 3 means that the actions planned in this group will start two years after the start of the European Collaborative Cloud for Cultural Heritage. The actions listed here are the ones who need a tight integration with the European Collaborative Cloud for Cultural Heritage services from the very beginning of their activities (thus, need an already quite consolidated status of the European Collaborative Cloud for Cultural Heritage).

Pilots 1-3: Interlinking collections and data production
Creating digital twins of collections of heritage artefacts in Europe. The focus is on stimulating digitization efforts, subject to a cultural scope (creation of a digital collection representing a specific

cultural context) and requiring the collaboration and joint work of several cultural institutions. The main focus could be on built heritage, archaeology, industrial heritage, etc.

Pilots 4-5: Experimenting and enhancing collaborative approaches
Experimenting with the European Collaborative Cloud for European Collaborative Cloud for Cultural Heritage collaborative and innovative uses of shared digital content (e.g., comparative analysis of artworks; innovative didactic experiences and related production of courseware). This might entail museum artworks, archaeology, architecture, local heritage, etc.

3.4 Governance, Business model and Sustainability

The initial question of the governance model is, if *one* exists at all. It is a cross-cutting issue guided both by technical choices and by the experiences of existing infrastructures including similar (Cloud) initiatives in other areas, which also meet the various objectives set by such an initiative of a European Collaborative Cloud for Cultural Heritage. Based on the previous recommendations - we have already mentioned in this report regarding the specific needs of Cultural Heritage communities and the technical choices that will result from it – that the governance we envisage is integrative and evolving. It must be able to take into account multi-level realities: from the local through the national up to European/international regulations, norms and standards.
Each governance model responds to specific objectives and characteristics of the participating actors, so it will be necessary to find the path that best embraces the specificity of the Cultural Heritage domain.

Thus, our approach is to study existing governance schemes, both close and distant to the Cultural Heritage domain, in order to analyze their strengths and weaknesses. We choose this approach, since it is not our role to propose a complex, ready-to-use scheme. On the contrary, we believe that it is essential to co-build following some clear guidelines. Furthermore, as we show below, it seems crucial to us to involve the main actors (defined in Part 1) from the outset, taking into account the diversity of scales, means, national and local realities, interwoven with the multi-level regulations.

The tasks of the European Collaborative Cloud for Cultural Heritage governing body will include:
- governing the European Collaborative Cloud for Cultural Heritage evolution;
- defining clear access and participation rules;
- defining a sustainability policy.

The guidelines will be structured around six main requirements (see subsection 3.4.1.) to build such a variable geometry governance.

Thus, the question of establishing a business model is also crucial and it is one of the tasks that should be accomplished jointly by the basic European Collaborative Cloud for Cultural Heritage consortium and the CSA project (which should jointly draw a governance and a sustainability policy). This topic includes the financial aspects of sustainability, i.e., who will pay for the implementation and who will pay for future enhancements and maintenance, which must be considered from the very outset of the consortium.

Finally, the question of sustainability, which has been mentioned many times in the background of this report and here, takes up the main technical, financial and human elements that we consider important to guarantee the sustainable development of a cloud for Cultural Heritage. The European

Collaborative Cloud for Cultural Heritage governing bodies should therefore deal with the issue of sustainability from three angles: technical, human and financial.

3.4.1 Proposed requirements for establishing a Governance schema

Ensuring the production and the management of data in the Cloud

At first sight, the envisioned governance must drive the technical robustness, management and security of services and tools. These features must be designed and implemented according to a general data strategy encompassing acquisition, organization, analysis, and delivery of data in support of community objectives. The common base for the European Collaborative Cloud for Cultural Heritage must be created technically on the basis of the needs expressed by the end-users, with common and shared data models and protocols. This strong "core" technical layer is essential to guarantee security and quality of data and processes supported by the European Collaborative Cloud for Cultural Heritage. By way of comparison, this is somewhat different from the approach taken in the EOSC, which aggregates useful services in all areas of science, by a disciplinary cluster from which cultural heritage as a field in its own right is absent, in order to fulfill the general objectives of open science.

Therefore, in terms of participation and access rights, any structure (museum or cultural institutions) or initiative wishing to participate (by using, even partially, a service; or by proposing a service or a tool, etc.), should be authorized to do so. But within this framework and in order to guarantee the possibility of integration while ensuring compliance and quality, a process of evaluation of services and tools to integrate the European Collaborative Cloud for Cultural Heritage infrastructure is necessary. That is why we suggest including independent expert bodies within the governance ecosystem. In addition, the services integrated in the subsequent phases of the project will certainly be subject to SLAs (Service Level Agreements) between the suppliers and the "core" consortium of European Collaborative Cloud for Cultural Heritage.

This type of organization should guarantee operational autonomy allowing any initiative to set up and develop its own projects and activities performed with the aid of the European Collaborative Cloud for Cultural Heritage tools and resources. Specific regulations should be defined to evaluate and grant authorization to participate to any potential Cultural Heritage user.

At the technical level, this translates into the recommendation of a SaaS approach (initial cloud architecture and core services), but open to PaaS to complete and adapt the user-oriented offer.

Choosing a proper variable-geometry governance schema

Observing existing models that have already started to evolve is a useful source of inspiration:

- At a time when European research infrastructures are taking stock, the positioning of the two most important ones dedicated to SSH, including Arts and Cultural Heritage, is useful for a future European Collaborative Cloud for Cultural Heritage dedicated to cultural institutions: ERICs that will partly integrate the EOSC, or even some existing e-infrastructure services (i.e., OpenAire or EGI);
- CLARIN is an intriguing model from the point of view of the technical integration of services offered to all and heavily "normed" and controlled to correspond to the standards of open science and now FAIR principles;
- The progressive but increasingly effective integration of the needs and productions of national communities at European level, i.e., the whole community building part, is one of the strong points of the ERIC DARIAH.

The choice of a European Collaborative Cloud for Cultural Heritage must take into account these points of success and difficulties in its own governance. Moreover, these European research infrastructures already have a strong base in the communities of which a part of the researchers associated with heritage sciences are aware. This can also constitute one of the vectors of knowledge produced for and by researchers thanks to the use of Cloud Cultural Heritage.

A fairly classic way of organizing the governance of this type of structure is coordination via national nodes or "hubs". These seem relevant here too, except that the national nodes should be allowed to adapt their own governance according to the reality on the ground, the organization of the actors in their sector in their area. One of the functions that could be assigned to these national nodes could be to coordinate the ingestion of identified big collections of digital Cultural Heritage objects or making the connection with existing national and/or international aggregators (e.g., Museum-Digital) in order to allow the development of an AI-based cloud services with them (e.g., semantic tagging).
In addition to the national nodes, transversal working groups in charge of developing the Cloud from a technical, human (see community building topic below), institutional and scientific point of view could also be created.

If we take the example of museums, each country has strategies for grouping, organizing and distributing responsibilities and competences internally, and often even with this, grouping strategies are created locally, autonomously and locally in the museums. Therefore, these realities on the ground must be taken into account when establishing the governance of a European Collaborative Cloud for Cultural Heritage and it is essential to think of an ecosystem and governance with variable geometry.

Independent assessment based on the stakeholders' needs is essential. Being based on an objective analysis and metrics of the use of the European Collaborative Cloud for Cultural Heritage data by an independent body, it should be the driving force of the future cloud adaptation and updates, as already mentioned in section 3.1 (https://www.cs.upc.edu/~pere/Long_term_adaptation_3.jpg). By including users and a guaranteed independent assessment, it should provide long-time benefit for the European Cultural Heritage community.

Ensuring the creation of an integrative community of Cultural Heritage professionals at various levels (scalability)

As we mentioned above, the evolutionary characteristic of the European Collaborative Cloud for Cultural Heritage and its community implies establishing an ingestion strategy, i.e., ensuring the cooperation of the technical and personal involvement of different types of actors, who work at different levels.
At the technical level, it is necessary to create the conditions for a dialogue with existing infrastructures, in particular those listed in Table 1, in order to establish strong links (aggregators, services), to provide the integration of services, when they correspond to the identified needs and to continue to exchange information. This should be part of a Call dedicated to consolidate effective governance for the Cloud.
In this perspective, it will be necessary to find proper contacts with key projects and infrastructures in order to establish cooperation with them and eventually to integrate them into the governance schema. A "governing board" could be installed with representatives from the field. This board could oversee and manage the lasting connection between the Cloud and its users to ensure commitment and participation.
In the same vein, to continue with the idea of dynamic inclusiveness, governance should not be restricted to the consortium that will be responsible for the basic European Collaborative Cloud for Cultural Heritage system. Governance should include a number of representative stakeholders (to be open to the voices of the GLAM community). Since we are envisioning the joint and coordinated contribution of multiple consortia (each one in charge of implementing some added functionalities of tools on top of the basic European Collaborative Cloud for Cultural Heritage), the European Collaborative Cloud for Cultural Heritage governance should also include representatives of these

other communities. Thus, it should be designed as an evolving body, able to include new representatives and open to the contribution of a wide community.
Finally, in order to mobilize a wide community and to make it cohesive over the years, we propose to integrate a peer review to increase scholarly research participation.

A system of ambassadors, appointed for one or two years, is an efficient option, which has been currently adopted in several organizations. Ambassadors increase the diversity of the group of actors. After being properly trained, ambassadors assist and advise museums and cultural institutions to start work in and with the Cloud locally, to discover its functionalities, thus contributing to the outreach, i.e., “evangelizing” the Cloud’s target audience.

The overall policy should also consider how and to what extent this European effort could be open to the rest of the world (including the Global South).

Ensuring the networking

Integrating this diversity of actors into the strategic governance of the Cloud will make it possible to identify the points of improvement in the relations between small and large museums and to encourage exchanges that are not, in the end, the daily business of actors in the same field.
The governance and functionalities of the European Collaborative Cloud for Cultural Heritage should facilitate the exchanges between large and small cultural institutions. If large museums have sufficient resources to process their collections and to carry out exchanges with other large structures, the benefits will be mutual. A better mutual knowledge will allow the discovery of unknown or forgotten pieces, poorly referenced. A work of cross-referencing the skills and expertise of the staff of cultural institutions (in collaboration for example with the 4CH project) could also benefit the large structures. Finally, working together within the framework of the same reference system, pooling resources and ideas to explore innovative pricing systems or new methods offered by advanced technologies is an added value that is difficult to foresee within the framework of a single local policy, or even a state policy.
The existence of these networks of European professionals and experts (including researchers) will also contribute to an enhanced visibility of the European know-how in the Cultural Heritage domain.

Proposing a complete training program

Alongside initiatives such as the ambassadors’ program, and schemes to encourage and recognize skills and good use of European Collaborative Cloud for Cultural Heritage services (see subsection 3.2.3), it will also be crucial to develop training materials and programs to raise skills. This need for training is an ongoing effort to be pursued during all phases of the Cloud development. This should entail creation of proper documentation and the organization of courses (face-to-face or distance learning) with dedicated materials and persons. This could give rise to "train the trainers" programs that consolidate a network of experts capable of providing effective and often local support, or the formation of a traveling team to help contributing museums (e.g., CLARIN to support the "opening" of CLARIN Centres).
As the European Collaborative Cloud for Cultural Heritage will grow and expand the scope of its users (and certainly of its functionality), specific programs should be dedicated to create content for training and education on digital technologies applied to Cultural Heritage. This kind of material could also be disseminated through existing infrastructures like DARIAH or E-RIHS (this way they will also be an input to EOSC).

Ensuring an economic and technical sustainability

The question of economic sustainability refers to both the maintenance of the services over time (under the monitoring of a group of independent experts as we have already mentioned) and to the capacity to provide a long-term preservation service for the data ingested in the European Collaborative Cloud for Cultural Heritage, and to integrate new perspectives towards a digital economy. Therefore, a side effect of the project could be to propose innovative business models properly adapted to the data and operating methods of museums and cultural institutions. It will also raise the issue of creating a business model for pricing certain services, or even the reuse of certain data.

3.4.2 Towards business models - Creating traceable "added values" of digital assets

The creation of a European Collaborative Cloud for Cultural Heritage opens a new scenario that could originate the setup of new or adapted business models. We start from the reasons and motivations for building a European Collaborative Cloud for Cultural Heritage (see section 1.1.). To evaluate possible business models, it is indeed necessary to focus on 1) why museums and cultural institutions would use such an infrastructure and 2) why they would contribute to the population of the Cloud (especially in the case of institutions which already have some basic internal infrastructure).

1) This question finds an easy answer: every potential service offered to professionals could be in principle useful and adopted. However, it will be assessed against cost and quality (efficiency and effectiveness).
2) The second question generates more difficulty to answer. A first point is technical: ingesting data or linking data from an external archive or repository should not be an excessively expensive action (considering human intervention). A second point is political: institutions should be assured that they do not lose control over their data, i.e., IPR will be respected and ownership of the data will be clearly acknowledged. Third point is the possibility of creating value from the heritage data housed in museums. European Collaborative Cloud for Cultural Heritage could offer to museums services for commercialization, such as links to creative industries that need access to high-quality raw digital cultural material. While for some structures the services offered will respond almost naturally to strongly identified needs, for others it will be necessary to persuade and to insist on the benefits of networking and also that of associated services and innovative projects. The security aspect is also essential in the choice of a European infrastructure whose governance will have to be variable geometry as we have specified, which should reassure the partners, in particular on their level of autonomy.

Improving collaborative documentation of heritage objects

A great number of studies (monographic and comparative investigations, etc.) are carried out on Cultural Heritage objects by students (from the first training cycles to the doctorate) and by researchers from different disciplines (history, archaeology, anthropology, sociology, geography, etc.) through a variety of partnerships and at different levels (universities, laboratories, national cultural institutions, museums and sites of local authorities, cultural associations, etc.). If the results of this work are systematically translated into "digital products" (study papers, reports, dissertations, scientific papers, surveying, hypothetical reconstructions, etc.), a *cloud platform* for the collection, inventory, exchange and multi-criteria research dedicated to the valorization of these resources would make visible a capital of knowledge that partnerships between research organizations and cultural institutions produce on a daily basis (from the scale of a single cultural object to that of a museum collection or even a corpus of studies related to the scale of the territory). This aspect can be linked to the question of the economic value of the resources able to describe, interpret and make understandable a Cultural Heritage object.

Moreover, all the cooperative instruments provided by the European Collaborative Cloud for Cultural Heritage will allow not only to document the results produced by those studies, but also to preserve the methodological path followed to reach them, contributing to an increased reproducibility of scientific and professional experiences. This will be an important added value for students, researchers and practitioners (e.g., for collecting guidelines and good practices in several application contexts).

A second key aspect of improving collective knowledge about cultural assets is the semantic enrichment of the digital resources that represent them.
Strategies and methods for the collaborative semantic enrichment can be integrated into the production chain of digital resources, by encompassing the entire data life cycle and by linking features of Cultural Heritage objects to the different gazes (disciplinary profiles, sensibilities, levels of interpretation, etc.), which are mobilized for their study, preservation and dissemination. Hence, it is a matter of using technologies (voice recognition, web semantics, VR /AR, etc.) not only to 'better present objects in digital environments', but to interweave the tangible and intangible dimensions of a Cultural Heritage object by closely linking the representation of the material object with the representation of the meaning of that object in different knowledge domains.

The two aspects presented (collaborative documentation and semantic enrichment) can produce an impact on digital economy themes, in particular stimulating the setup of mechanisms that museums are currently unable to develop and adopt autonomously. An important objective would be here to enable the traceability of a value production chain, covering the entire life cycle: from the creation of a digital resource (taking into account the technical but also and above all the intellectual contributions), to the multiple semantic enrichment steps through collaborative (and even participatory) scenarios. This would make it possible to anchor the memory of the (technical and intellectual) production chain within the digital cultural assets. Blockchain-like approaches and new frameworks for the monetization of digital rarities (e.g., NFTs - non-fungible tokens) could be explored in order to translate the traceability of a value production chain into potential opportunities to build new business models.

Financial sustainability

It is crucial to find a proper framework for supporting financially the development and maintenance of software including services and tools. The establishment of such a framework should satisfy the following queries regarding the medium and long-term financing of a European Collaborative Cloud for Cultural Heritage:

- *Who pays for the implementation?* The initial implementation (the basic European Collaborative Cloud for Cultural Heritage and the further projects which will extend it) should be based on EC funds (since we cannot base these actions on private funds or National investments), based on competitive calls. This will be the seed money that will allow to ignite this activity (timely and with the needed public funding). On this point, and to bridge the gap to the next, the Member States must play a key role. In addition to the European funding that will serve as a springboard for the construction of this Cloud, they appear to be in the best position to ensure its funding, as is the case, more or less directly, for other initiatives of the same type in other fields: GAIA-X for companies and institutions, EOSC as we have already mentioned (particularly through the ERICs) or eHDSI for the very sensitive field of health data (https://ec.europa.eu/health/ehealth-digital-health-and-care/electronic-cross-border-health-services_en).
- *Who pays for future enhancements and maintenance?* This is also a critical point. Keeping the European Collaborative Cloud for Cultural Heritage alive and operational after the end of the first group of projects requires (smaller) public funding. This funding should be assigned on the base of a timely and precise assessment of the effectiveness of the various components

(evolutionary model), to guarantee the maintenance and extension of the tools which have been granted a strong support and use by the community. This is a radical policy change in the way EC maintains the investment; according to our vision, the technical resources included in the European Collaborative Cloud for Cultural Heritage should enable the EC and the European Collaborative Cloud for Cultural Heritage governance in making choices which would be not political but, again, usage-driven. Then, it is also possible to foresee that some tools can be paid additions, or that maybe support could drive revenue. Here, we can refer to an example of the infrastructure widely used by small museums: https://en.about.museum-digital.org/. In Hungary, a company https://muzeumdigitar.hu/ provides paid support, goes to the museums and convinces them, and although the cloud service is free, their “help” is costly, it is still successful.

- *Only public funding?* The pay-per-use model is not adequate to an under-funded community (consider small and medium museums and cultural institutions). However, the material created and stored on the European Collaborative Cloud for Cultural Heritage could be of interest for commercial purposes (creative industries, publishers, video production, tourism). As mentioned above, re-use of data could produce revenues. One side mission of the European Collaborative Cloud for Cultural Heritage could be to create an organization and a support for selling digital assets (in a controlled manner) and also to redirect shares of the revenues to the original owners (of both the original artwork and of the digital counterpart).

3.4.3 A multi-faceted sustainability

Long-term preservation vs archiving

Long-term digital preservation (LTDP) means keeping digital information understandable and usable for several decades and even hundreds of years, even as hardware, software, and file formats, for example, become obsolete and change during this time. On the other hand, the fact of archiving is a matter of political choices which it is not for us to analyze here. The two activities are closely linked, since in order to make archival choices and intend to preserve archived digital data, it is necessary to consider the conditions for long-term preservation of the data. Thus, it could be said that long-term preservation is a necessary condition for archiving. This concern for long-term preservation, applied to digital twins for example, is all the more crucial today as some pieces of European Cultural Heritage may be threatened in their physical existence (natural disasters, political unrests, etc.).

Reliable long-term preservation requires active monitoring of content integrity and preparedness for a wide range of risks. Metadata describing, for example, the content, history and origin of the material and information about the technicalities of its use must be provided. Finally, for long-term preservation of data, it is necessary to apply the FAIR principles which partly guarantee its quality (completeness of metadata, PIDs, etc.). At this stage, it may also be appropriate to jointly build a tool that is able to identify the formats and standards accepted by the Cultural Heritage community, but also to measure the quality of the submitted files. This tool could be backed up by the first brick of the Cloud, "Data" (see Figure 2). This process has been tested in France for almost twenty years.

The European Collaborative Cloud for Cultural Heritage should include an LTDP service that is intended for the long-term storage of selected data in the Cloud. It would complement national and other LTDP services. This enables good management, secure storage, and findability of research data and related metadata. Ownership of the materials remains with the organizations throughout the retention period. An organization that has exported material to the LTDP service can retrieve the material back to itself. The LTDP service is primarily a service focused on the storage of data.

A reference model at the European level for this could be the Zenodo service [94] that is targeted for storing scientific datasets. Possibly this ready-to-use service could be extended for Cultural Heritage data. There are already Cultural Heritage datasets stored in Zenodo, such as the Mapping Manuscripts Migrations Knowledge Graph data about mediaeval and Renaissance manuscripts from U.K., U.S., and France [95].

The sustainability of the infrastructure (and associated data) that we propose must imperatively be connected to a political vision of privileging sovereignty over European Cultural Heritage data, then, technically, to robust, trusted infrastructures, in line with the European research infrastructure roadmap and with duly identified infrastructures at the national level, which already display a variety of archival services and policies. As noted above, the notion of long-term preservation is different from the act of archiving, which remains, a priori, a national prerogative.
The French example and the collaboration between the national infrastructures Huma-Num (SSH), the CINES (National Computing Centre for Higher Education) and the National Archives offers a notable case to study these aspects.

Feedbacks at the national level

Various feedbacks at national level allow to measure practices and to confront regulations in view of a scaling up, or at least of a coordination between a European Collaborative Cloud for Cultural Heritage system for Cultural Heritage and services and/or institutions dedicated to preservation and archiving. From this point of view, it is compelling to relate successful experiences, as is the case in France with the collaboration between Huma-Num (https://www.huma-num.fr/), the French infrastructure for SSH and CINES, the French National Computing Centre for Higher Education (https://www.cines.fr/en/). This is how the partnership, which has now been in existence for almost ten years and governed by a four-year agreement, came about and benefited France.
The French infrastructure Huma-Num offers to the community of producers of digital data in SSH a long-term preservation service. But in this offer, Huma-Num does not directly operate the service, it rather supports teams and research projects that request it, until their data are actually deposited at CINES for the long term. It relies, for this activity, on the infrastructure and skills of a certified centre, the CINES. For its part, the CINES remains in its role of providing preservation of data and digital documents produced by the French community of Higher Education and Research. It offers digital archiving solutions for medium- and long-term preservation, solutions which are shared and customizable which allows a real collaboration with Huma-Num. For example, this has allowed Huma-Num to conduct a study in collaboration with experts from the TEI community to introduce this format at CINES, while it was not previously supported. The level of requirement defined for the integration of this format at CINES has allowed in return to all producers using this format to improve the quality of their production, in particular their structuring and documentation. It was the same for a work provided in collaboration with the Consortium Huma-Num 3D for Humanities (https://shs3d.hypotheses.org/) around a 3D data archiving format (PLY and Collada), now supported by the CINES. For more information about this service in partnership, see https://documentation.huma-num.fr/humanum-en/#long-term-preservation.
In the field of Cultural Heritage, some big cultural institutions, like the French National Library (BNF: https://www.bnf.fr/fr/spar-systeme-de-preservation-et-darchivage-reparti) have already developed their own Preservation System. In that case, it will be important to take into account the interoperability and services accessibility in these cases.

Sustainability of usage

"Sustainability of usage" highlights all the human aspects behind the term infrastructure. Finally, the question of the evolutionary aspect of the Cloud and the sustainability of its use and its development over time will depend both on the ease of access to the proposed services and on the training associated with the use of these services, which should not be limited to accessible documentation but to a real human infrastructure. This is why we have already mentioned the need for a system of user fidelity in subsection 3.2.3, but also the importance of consolidating an inclusive and dynamic community, representative of the diversity of the target users, and keeping it trained, throughout the development of the project and in the long term, as it is indicated among the main requirements in subsection 3.4.1.

3.5 Contributions to Green Deal

Though Green Deal transformation of our society is not the primary focus of the action described in this report, but it is worthwhile to spend a few words locating also our Cultural Heritage-focused actions in the bigger context of the climate-aware actions and to discuss briefly the potential impact of the actions proposed to the state of our planet and society.

First, we know that any digitization effort, such as the creation of a European Collaborative Cloud for Cultural Heritage, brings new needs for energy consumption, thus producing an impact over our climate. Data management has been already envisaged as a critical consumer of electricity worldwide (various analyses suggest that data centres represent 1-2 percent of global electricity consumption). The European Collaborative Cloud for Cultural Heritage will not be a counter-example, it will increase electricity consumption, especially if it will become a very successful resource (storage and access figures could become impressively high). On the other hand, having a European Collaborative Cloud for Cultural Heritage will allow us to move many existing data and processing to more energy-efficient technologies. The European Collaborative Cloud for Cultural Heritage should replace a plethora of small computers, which have or might be installed to enable local hosting of data; these resources would be (at least partially) replaced with a professional cloud infrastructure, managed following sustainable criteria (adopting green computing technologies, using sophisticated thermal power to manage heating, etc). If we consider the sum of new consumption and reduced consumption (due to the eliminated local storages) the overall impact of the European Collaborative Cloud for Cultural Heritage could therefore be even neutral.

A smaller impact will be the reduction of travels related to the possibility of studying and visiting places and inspecting artworks at a distance, using the digital clones. This will reduce the need to move to the place where the Cultural Heritage asset is conserved, will support or facilitate a number of activities which can be done at a distance, based on the use of digital clones. There is not the risk of reducing substantially the personal experiences on the real place (e.g., visiting a museum or a historical city), since the real world will maintain forever an aura and a potential experience that cannot be replaced with digital instruments (thus, we do not foresee a potential negative impact over tourism and related economy). Conversely, web instruments enabling the discovery of Cultural Heritage assets located in remote locations (small museums, small cities) will help relocate tourist flow from first-rank and crowded destinations to less-crowded targets.

The European Collaborative Cloud for Cultural Heritage will also contribute to the **digital transformation policies** and related implementation paths. Cultural Heritage is a rather conservative domain, a major investment in a European Collaborative Cloud for Cultural Heritage will support and popularize many new instruments for moving study and conservation activities in the digital domain. A considerable impact could be also the use of these digital assets to support teaching of art- and Cultural Heritage-related matters in any level of our education system.

REFERENCES

[1] IaaS vs PaaS vs SaaS, article in RedHat Magazine, https://www.redhat.com/en/topics/cloud-computing/iaas-vs-paas-vs-saas

[2] John J. Godfrey, Antonio Zampolli, "Language resources", Chapter from "Survey of the state of the art in human language technology", 1997, pp.381-384.

[3] Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., & Chamzas, C. (2007). Methods for 3D digitization of Cultural Heritage. Journal of Cultural Heritage, 8(1), 93–98. https://doi.org/10.1016/j.culher.2006.10.007

[4] Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., & James, S. (2020). Machine Learning for Cultural Heritage: A Survey. Pattern Recognition Letters, 133, 102–108. https://doi.org/10.1016/j.patrec.2020.02.017

[5] Dell'unto N., Leander A., Dellepiane M., Callieri M., Ferdani D., Lindgren S., 2013. Digital reconstruction and visualisation in archaeology: Case-study drawn from the work of the Swedish Pompeii Project, Proceedings of the 2013 Digital Heritage International Congress, Marseille, France, IEEE, 28 oct - 01 nov 2013, p. 621-628. DOI : 10.1109/DigitalHeritage.2013.6743804.

[6] FERDANI, Daniele, DEMETRESCU, Emanuel, CAVALIERI, Marco, PACE, Gloriana et LENZI, Sara, 2020. 3D Modelling and Visualisation in Field Archaeology. From Survey To Interpretation Of The Past Using Digital Technologies, Groma. Documenting archaeology, 10 April 2020, Vol. 15. DOI : 10.12977/groma26.

[7] DEVLIN, Kate, CHALMERS, Alan et BROWN, Duncan, 2003. Predictive lighting and perception in archaeological representations, UNESCO World Heritage in the Digital Age: 30th Anniversary Digital Congress, Vol. 14.

[8] Veronica, CHALMERS, Alan et MARTINEZ, Philippe, 2004. High fidelity reconstruction of the ancient Egyptian temple of Kalabsha, Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa - AFRIGRAPH 2004, ACM Press, p. 107-115.

[9] Hoffmeister, Dirk, 2017. Simulation of tallow lamp light within the 3D model of the Ardales Cave, Spain, Quaternary International, Feb. 2017, Vol. 430, p. 22-29. DOI : 10.1016/j.quaint.2016.05.010.

[10] BOURDIER, Camille, FUENTES, Oscar et PINÇON, Geneviève, 2015. Contribution of 3D technologies to the analysis of form in late palaeolithic rock carvings: The case of the Roc-aux- Sorciers rock-shelter (Angles-sur-l'Anglin, France), Digital Applications in Archaeology and Cultural Heritage, Vol. 2, n° 2-3, p. 140-154. DOI : 10.1016/j.daach.2015.05.001.

[11] Vallet, J.-M., De Luca, L., Feillou, M., Guillon, O., & Pierrot-Deseilligny, M. (2012). An Interactive 3-Dimensional Database Applied to the Conservation of a Painted Chapel. International Journal of Heritage in the Digital Era, 1(2), 233–250. https://doi.org/10.1260/2047-4970.1.2.233

[12] Apollonio, F., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A., Scopigno, R., & Sobrà, G. (2017). A 3D-centered information system for the documentation of a complex restoration intervention. Journal of Cultural Heritage, 29, 89-99.

[13] McCarthy, J. (2014). Multi-image photogrammetry as a practical tool for cultural heritage survey and community engagement. Journal of Archaeological Science, 43, 175–185. https://doi.org/10.1016/j.jas.2014.01.010

[14] Kirchhöfer, M., Chandler, J., & Wackrow, R. (2011). Cultural Heritage Recording Utilising Low-Cost Closerange Photogrammetry. Geoinformatics FCE CTU, 6, 185–192. https://doi.org/10.14311/gi.6.24

[15] Ridge, M., Blickhan, S., Ferriter, M., Mast, A., Brumfield, B., Wilkins, B., … Prytz, Y. B. (2021). 3. Why work with crowdsourcing in cultural heritage? In The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage - community review version. Retrieved from https://britishlibrary.pubpub.org/pub/why-work-with-crowdsourcing-in-cultural-heritage

[16] DOERR, Martin, TZOMPANAKI, Katerina, THEODORIDOU, Maria, GEORGIS, Christos, AXARIDOU, Anastasia et HAVEMANN, Sven, 2010. A Repository for 3D Model Production and Interpretation in Culture and Beyond, VAST: International Symposium on Virtual Reality, 8 p. DOI : 10.2312/VAST/VAST10/097-104.

[17] N Lercari, D Jaffke, A Campiani, A Guillem, S McAvoy, et al, Building Cultural Heritage Resilience through Remote Sensing: An Integrated Approach Using Multi-Temporal Site Monitoring, Datafication, and Web-GL Visualization, Remote Sensing, 2021.

[18] Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavrilis, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C., & Vlachidis, A. (2017). ARIADNE. Journal on Computing and Cultural Heritage, 10(3), 1–27. https://doi.org/10.1145/3064527

[19] DUDEK, I. et BLAISE, J. Y., 2017. What Comes before a Digital Output? Eliciting and Documenting Cultural Heritage Research Processes, International Journal of Culture and History (EJournal), Vol. 3, n° 1, p. 86-97. DOI : 10.18178/ijch.2017.3.1.083.

[20] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies, 43(5–6), 907–928. https://doi.org/10.1006/ijhc.1995.1081

[21] DOERR, Martin, 2005. The CIDOC CRM, an Ontological Approach to Schema Heterogeneity, Semantic Interoperability and Integration. Dagstuhl, Germany : Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

[22] PHILLIPS, Stephen, WALLAND, Paul, MODAFFERI, Stefano, DORST, Leo, SPAGNUOLO, Michela, CATALANO, Chiara Eva, OLDMAN, Dominic, TAL, Ayellet, SHIMSHONI, Ilan et HERMON, Sorin, 2016. GRAVITATE: Geometric and Semantic Matching for Cultural Heritage Artefacts, Eurographics Workshop on Graphics and Cultural Heritage, 4 p. DOI : 10.2312/GCH.20161407.

[23] OLDMAN, Dominic and TANASE, Diana, 2018. Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace, The Semantic Web – ISWC 2018, Springer International Publishing, Lecture Notes in Computer Science, p. 325-340. ISBN : 978-3-030- 00667-9.

[24] Baca, M. (2003). Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information. Cataloging & Classification Quarterly, 36(3–4), 47–55. https://doi.org/10.1300/j104v36n03_05

[25] Doulamis, A., Doulamis, N., Protopapadakis, E., Voulodimos, A., & Ioannides, M. (2018). 4D Modelling in Cultural Heritage. In Lecture Notes in Computer Science (pp. 174–196). Springer International Publishing. https://doi.org/10.1007/978-3-319-75789-6_13

[26] Havemann, S., Settgast, V., Berndt, R., Eide, Ø., & Fellner, D. W. (2009). The arrigo showcase reloaded—towards a sustainable link between 3D and semantics. Journal on Computing and Cultural Heritage, 2(1), 1–13. https://doi.org/10.1145/1551676.1551680

[27] Oreni, D. (2013). From 3D Content Models to HBIM for Conservation and Management of Built Heritage. In Lecture Notes in Computer Science (pp. 344–357). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39649-6_25

[28] Luczfalvy Jancsó, A., Jonlet, B., Hoffsummer, P., Delye, E., & Billen, R. (2017). An Analytical Framework for Classifying Software Tools and Systems Dealing with Cultural Heritage Spatio-Temporal Information. In Lecture Notes in Geoinformation and Cartography (pp. 325–337). Springer International Publishing. https://doi.org/10.1007/978-3-319-63946-8_50

[29] Tangelder, J. W. H., & Veltkamp, R. C. (n.d.). A survey of content based 3D shape retrieval methods. Proceedings Shape Modeling Applications, 2004. Proceedings Shape Modeling Applications, 2004. https://doi.org/10.1109/smi.2004.1314502

[30] Koller, D., Frischer, B., & Humphreys, G. (2009). Research challenges for digital archives of 3D cultural heritage models. Journal on Computing and Cultural Heritage, 2(3), 1–17. https://doi.org/10.1145/1658346.1658347

[31] Samoun, L., Fisichella, T., Lingrand, D., Malleus, L., & Precioso, F. (2018, October). An Interactive Content-Based 3D Shape Retrieval System for on-Site Cultural Heritage Analysis. 2018 25th IEEE International Conference on Image Processing (ICIP). 2018 25th IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip.2018.8451546

[32] Erik Champion and Hafizur Rahaman (2020), "Survey of 3D Digital Heritage Repositories and Platforms", Virtual Archaeology Review, Vol. 11 (23), pp. 1-15, 2020: https://doi.org/10.4995/var.2020.13226

[33] Marco Fioruccia, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, Stuart James (2020), "Machine Learning for Cultural Heritage: A Survey", Pattern Recognition Letters Vol. 133 (2020), pp. 102–108: https://www.sciencedirect.com/science/article/pii/S0167865520300532

[34] Kyriacos Themistocleous (2020), "The Use of UAVs for Cultural Heritage and Archaeology"; In: Hadjimitsis D. et al. (Eds) Remote Sensing for Archaeology and Cultural Landscapes. Springer Remote Sensing/Photogrammetry. Springer, Cham: https://doi.org/10.1007/978-3-030-10979-0_14

[35] Hussain A.J., Al-Fayadh A., Radi N.,Image compression techniques: A survey in lossless and lossy algorithms, Neurocomputing, Volume 300, Pages 44-69 (2018). ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2018.02.094.

[36] Google Maps (http://maps.google.com) and Google Earth (https://www.google.it/intl/it/earth/)

[37] Potdar V. M., Han S. and Chang E., "A survey of digital image watermarking techniques," *INDIN '05. 2005 3rd IEEE International Conference on Industrial Informatics,* Perth, WA, Australia, pp. 709-716 (2005). doi: 10.1109/INDIN.2005.1560462.

[38] M. Picollo, C. Cucci, A Casini, L Stefani , "Hyper-spectral imaging technique in the cultural heritage field: new possible scenarios", Sensors, 2020, *20*(10), 2843; https://doi.org/10.3390/s20102843

[39] Malzbender T., Gelb D., Wolters H., Polynomial texture maps. In: Proc. of the 28th ACM SIGGRAPH Conference, pp. 519–528. ACM (2001).

[40]https://en.wikipedia.org/wiki/360_photography ; for an example of panoramic images adopted to enable the virtual visit to museum see also: http://www.youvisit.com/tour/louvremuseum (accessed on 26 November 2020). As a second example, see: M. Comino, C. Andujar, A. Chica, and P. Brunet. "Error-aware construction and rendering of multi-scan panoramas from massive point clouds". In: Computer Vision and Image Understanding 157 (2017). Large-Scale 3D Modeling of Urban Indoor or Outdoor Scenes from Images and Range Scans, pp. 43–54

[41] Marco Callieri, Matteo Dellepiane, Paolo Cignoni, Roberto Scopigno, "Processing sampled 3D data: reconstruction and visualization technologies", in "Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration and Reconstruction of Ancient Artworks", Taylor and Francis, page 105—136, 2011.

[42] Sketchfab, a 3D web viewer (https://sketchfab.com/)

[43] Roberto Scopigno, "Mixing Visual Media for Cultural Heritage", in Emerging Technologies and the Digital Transformation of Museums and Heritage Sites, First International Conference, RISE IMET 2021, Cyprus, June 2–4, 2021, pp. 297-315

[44] Pere Brunet, Carlos Andujar, "Immersive data comprehension: visualizing uncertainty in measurable models", Frontiers in Robotics and AI, 2015, doi: 10.3389/frobt.2015.00022. Available online: https://www.frontiersin.org/articles/10.3389/frobt.2015.00022/full

[45] Guidi, G. (2013). "Metrological characterization of 3D imaging devices," in Proceedings of the XII SPIE Optical Metrology Videometrics, Range Imaging, and Applications Conference (Munich), 87–91. doi:10.1117/12.2021037

[46] JCGM (2008) "Evaluation of measurement data. guide to the expression of uncertainty in measurement". Working Group 1 of the Joint Committee for Guides in Metrology (JCGM/WG 1). JCGM 100:2008. 1st Ed. Available at: http://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e_FILES/MAIN_C045315e/Start_e.html

[47] Pauly, M., Mitra, N. J., and Guibas, L. J. (2004). "Uncertainty and variability in point cloud surface data," in Proceedings of the First Eurographics Conference on Point-Based Graphics, SPBG'04. eds M. Alexa, M. Gross, H. Pfister, and S. Rusinkiewicz (Aire-la-Ville: Eurographics Association), 77–84.

[48] Domingue, J., Fensel, D., Hendler, J.A. (2011). Introduction to the semantic web technologies. In:Domingue, J., Fensel, D., Hendler, J.A. (eds.) Handbook of Semantic Web Technologies, pp.1–41. Springer (2011). https://doi.org/10.1007/978-3-540-92913-01

[49] Semantic Web standards: https://www.w3.org/standards/semanticweb/ Accessed 2021-01-16.

[50] Gutierrez, C., Sequeda, J.F. (2021). Knowledge graphs. Communications of the ACM 64 (3), 96–104 https://doi.org/10.1145/3418294

[51] Heath, T., Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space (1st edition).Morgan & Claypool, Palo Alto, California, http://linkeddatabook.com/editions/1.0/

[52] Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the Semantic Web. Morgan & Claypool, Palo Alto, California.

[53] Eva Pietroni and Daniele Ferdani (2021), "Virtual Restoration and Virtual Reconstruction in Cultural Heritage: Terminology, Methodologies, Visual Representation Techniques and Cognitive Models", Information 2021, Vol. 12, 167: https://doi.org/10.3390/ info12040167

[54] Marco Fioruccia, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, Stuart James (2020), "Machine Learning for Cultural Heritage: A Survey", Pattern Recognition Letters Vol. 133 (2020), pp. 102–108: https://www.sciencedirect.com/science/article/pii/S0167865520300532

[55] Eero Hyvönen, Ruth Ahnert, Sebastian E. Ahnert, Jouni Tuominen, Eetu Mäkelä, Miranda Lewis and Gertjan Filarski: Reconciling metadata. Reassembling the Republic of Letters in the Digital Age (H. Hotson and T. Wallnig (eds.)), pp. 223-235, Göttingen University Press, 2019. https://doi.org/10.17875/gup2019-1146

[56] https://www.json.org/, accessed 06/01/2022.

[57] https://www.w3.org/RDF/, accessed 06/01/2022.

[58] http://openrefine.org/, accessed 12/06/2018. – See below 4.4.

[59] http://usc-isi-i2.github.io/karma/, accessed 07/07/2018.

[60] Hyvönen, Eero. Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Morgan & Claypool, Palo Alto, California, 2012. https://doi.org/10.2200/S00452ED1V01Y201210WBE003

[61] http://dublincore.org/ , accessed 06/01/2022.

[62] Staab, Steffen and Studer, Rudi (Eds.). Handbook on Ontologies, 2nd edition. Springer-Verlag, 2009.

[63] CIDOC CRM. ICOM/CIDOC Documentation Standards Group / CIDOC CRM Special Interest Group: Definition of the CIDOC Conceptual Reference Model, Version 6.2.3. Ore, C.-E., Doerr, M., Le Bœuf, P., Stephen Stead, S. (eds.), 2018. http://www.cidoc-crm.org/Version/version-6.2.3

[64] Cf. FRBR. IFLA Study Group on the Functional Requirements for Bibliographic Records: Functional Requirements for Bibliographic Records, Revised 2009. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records

[65] https://www.ifla.org/publications/node/11412, accessed 06/01/2022.

[66] Bekiari, C., Doerr, M., Le Bœuf, P., and Riva, P. (eds.). IFLA Working Group on FRBR/CRM Dialogue: Definition of FRBRoo: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism, Version 2.4., 2016. https://www.ifla.org/publications/node/11240

[67] *Rao, Delip, McNamee, Paul, and Dredze, Mark. Entity Linking: Finding Extracted Entities in a Knowledge Base. In Multi-source, Multilingual Information Extraction and Summarization, Theory and Applications of Natural Language Processing, Springer-Verlag, 2013, 93–115.*

[68] Hachey, Ben; Radford, Will; Nothman, Joel; Honnibal, Matthew; Curran, James R.: Evaluating entity linking with Wikipedia. Artificial Intelligence, Vol. 194, 130–150, 2013.

[69] Eero Hyvönen: Preventing Interoperability Problems Instead of Solving Them. Semantic Web Journal, vol. 1, no. 1-2, pp. 33-37, December, 2010.

[70] Wang, Zeyu et al. “CHER-Ob: A Tool for Shared Analysis and Video Dissemination", ACM Journal on Computing and Cultural Heritage, vol 11, issue 4 (2018).

[71] Ponchio F., Potenziani M., Dellepiane M., Callieri M.; Scopigno R. (2016). ARIADNE Visual Media Service: easy web publishing of advanced visual media, in Campana, Scopigno (Ed.), *CAA 2015 Proceedings*, Archaeopress, 2016. http://vcg.isti.cnr.it/Publications/2015/PPDCS15/

[72] Potenziani M., Callieri M., Dellepiane M., Corsini M., Ponchio F., Scopigno R. (2015). 3DHOP: 3D Heritage Online Presenter. *Computer & Graphics*, Volume 52, page 129—141. http://vcg.isti.cnr.it/Publications/2015/PCDCPS15/

[73] Ponchio F., Dellepiane M., Callieri M., Scopigno R. "Effective Annotations Over 3D Models", Computer Graphics Forum, Volume 39, Issue 1, February 2020, pp. 89-105 https://doi.org/10.1111/cgf.13664

[74] Messaudi T., Veron P., Halin G., De Luca L., An ontological model for the reality-based 3D annotation of heritage building conservation state - Journal of Cultural Heritage - Vol. 29, p.100-112 (2018).

[75] F. Ponchio, M. Dellepiane, "Fast decompression for web-based view-dependent 3D rendering"

ACM Web3D 2015. Proceedings of the 20th International Conference on 3D Web Technology, pp. 199-207.

[76] Fauzia Albertin et al., "Ecce Homo by Antonello da Messina, from non-invasive investigations to data fusion and dissemination", Scientific Reports, Volume 11, Number 1, page 15868-15885 - 2021.

[77] S. Rizvic *et al*., "Guidelines for interactive digital storytelling presentations of cultural heritage," *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, 2017, pp. 253-259, doi: 10.1109/VS-GAMES.2017.8056610.

[78] Drossis G., Birliraki C., Stephanidis C. (2018), "Interaction with Immersive Cultural Heritage Environments Using Virtual Reality Technologies". In: Stephanidis C. (Eds) HCI International 2018 – Posters' Extended Abstracts. HCI 2018. Communications in Computer and Information Science, vol 852. Springer, Cham: https://doi.org/10.1007/978-3-319-92285-0_25

[79] Bekele Mafkereseb Kassahun, Champion Erik (2019), "A Comparison of Immersive Realities and Interaction Methods: Cultural Learning in Virtual Heritage", Frontiers in Robotics and AI, Vol. 6, 2019: https://www.frontiersin.org/article/10.3389/frobt.2019.00091

[80] Bekele, M.K. (2021), "Clouds-Based Collaborative and Multi-Modal Mixed Reality for Virtual Heritage", Heritage 2021, Vol. 4, pp. 1447-1459: https://doi.org/10.3390/heritage4030080

[81] C. Andujar, A. Chica, P. Brunet (2012), "User-interface design for the Ripoll Monastery exhibition at the National Art Museum of Catalonia", Computers & Graphics, Vol, 36 (1), 2012, pp. 28-37: https://doi.org/10.1016/j.cag.2011.10.005

[82] Andujar, C., Brunet, P., Buxareu, J., Fons, J., Laguarda, N., Pascual, J. and Pelechano, N. (2018), "VR-assisted architectural design in a heritage site: the Sagrada Família case study", in EUROGRAPHICS Workshop on Graphics and Cultural Heritage. "GCH 2018, Eurographics Workshop on Graphics and Cultural Heritage: Vienna, Austria, November 12-15, 2018". European Association for Computer Graphics (Eurographics), 2018, p. 47-56: https://upcommons.upc.edu/handle/2117/129963

[83] Egyptian Ministry of Antiquities & American Research Center of Egypt (2020): Virtual visit to the Tomb of Menna in the Theban Necropolis: https://my.matterport.com/show/?m=vLYoS66CWpk - Also, virtual visit to the Tomb of Queen Meresankh III: https://my.matterport.com/show/?m=d42fuVA21To

[84] Tunkelang, D. (2009). Faceted Search. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan-Claypool, Palo-Alto. https://doi.org/10.2200/S00190ED1V01Y200904ICR005

[85] Hyvönen, Eero. Using the semantic web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. Semantic Web – Interoperability, Usability, Applicability 11(1), pp. 187–193, 2020. https://doi.org/10.3233/sw-190386

[86] Use of CAD systems for the documentation of the conservation status of the Coliseum, Rome, Istituto Superiore per la Conservazione ed il Restauro (ISCR), Rome, 2010-2012.

[87] Apollonio F.I., Basilissi V., Callieri M., Dellepiane M., Gaiani M., Ponchio F., Rizzo F., Rubino A., Scopigno R., Sobrà G., A 3D-centered information system for the documentation of a complex restoration intervention, Journal of Cultural Heritage, Volume 29, page 89-99 (2017).

[88] R. Scopigno, "Mixing Visual Media for Cultural Heritage", in Emerging Technologies and the Digital Transformation of Museums and Heritage Sites (First International Conference, RISE IMET 2021, Nicosia, Cyprus, June 2–4, 2021), Springer, 2021, pp.297-315.

[89] P. Centorrino, A. Corbetta, E. Cristiani, E. Onofri, "Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization", Journal of Computational Science, Vol. 53, 2021, 101357, ISSN 1877-7503.

[90] K. NGAMAKEUR, S. YONGCHAREON, J.YU, S. UR REHMAN "A Survey on Device-free Indoor Localization and Tracking in the Multi-resident Environment", ACM Computing Surveys Volume 53, Issue 4, July 2021, 29 pages, ISSN:0360-0300.

[91] Marcia Zeng and Jian Qin. 2022. Metadata, Third Edition. ALA Neal-Schuman, Chicago.

[92] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information extraction meets the semantic web: A survey, Semantic Web – Interoperability, Usability, Applicability 11 (2020) 255–335.

[93] Eero Hyvönen, How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web. Keynote presentation at DCMI 2021 conference (Dublin Core Matadata Initiative). Paper: https://seco.cs.aalto.fi/publications/2022/hyvonen-infra-2022.pdf; video: https://vimeo.com/620820788

[94] Sonia Shahzadi (2017) "Infrastructure as a Service (IaaS): A Comparative Performance Analysis of Open-Source Cloud Platforms", 2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD): https://ieeexplore.ieee.org/abstract/document/8031522

[95] Ivan Voras, Marin Orlić, Branko Mihaljević (2012), "An Early Comparison of Commercial and Open-Source Cloud Platforms for Scientific Environments", Lecture Notes in Computer Science book series (LNCS, volume 7327): https://link.springer.com/chapter/10.1007/978-3-642-30947-2_20

[95] Yasir Saeed (2021), "Top 5 Open Source Cloud Storage Software in 2021", Containerize: https://blog.containerize.com/2021/06/25/top-5-open-source-cloud-storage-software-in-2021/

[96] How Nextcloud compares to these popular closed-source services: https://nextcloud.com/compare/

[97] Top 5 Free Open Source Cloud computing platform for File Sharing: https://contenteratechspace.com/blogs/top-5-free-open-source-cloud-file-sharing-platforms/

[98] Zenodo service: https://zenodo.org/ Accessed 2022-02-16.

[99] Mapping Manuscript Migrations Knowledge Graph stored in Zenodo: https://zenodo.org/record/4440464#.YeQYc_5Bzns) Accessed 2022-02-16.

ABOUT THE AUTHORS

Pere Brunet is a retired Full Professor in Computer Science at the Polytechnic University of Catalonia in Barcelona. He has been actively working in Computer Science since 1967. He started his research work in Computer Graphics in 1979, his group being a pioneer on Computer Graphics research in Spain. His research interests include the investigation of digital 3D representations, geometry processing and immersive interaction in fields including Cultural Heritage. He is a member of the Royal Academy of Engineering of Spain and a member of the Academia Europaea. He has received a number of Awards like the Distinguished Career Award of the Eurographics Association and the Spanish National Informatics Award "Jose Garcia Santesmases" in 2010. https://www.cs.upc.edu/~pere/

Livio De Luca (Architect, PhD in Engineering, HDR in Computer Science), is Director of Research at CNRS and Director of the CNRS/MC- MAP Unit. Co-chair of the international congress UNESCO /IEEE/EG DigitalHeritage (2013, 2015) and coordinator of national and international projects, his research activities focus on the acquisition, modelling and semantic enrichment of multidimensional representations of cultural heritage objects. He's currently editor of the Journal of Cultural Heritage (Elsevier) and, since 2019, coordinates the Digital Data Working Group of the CNRS/Ministry of Culture scientific site for the restoration of Notre-Dame de Paris. http://www.map.cnrs.fr/ldl/

Eero Hyvönen is professor of semantic media technology at the Aalto University, Dept. of Computer Science, and director of Helsinki Centre for Digital Humanities (HELDIG) at the University of Helsinki. Since 2001, a major recent theme in his research has been development of the national level semantic web infrastructure in Finland and its applications especially related to the Cultural Heritage domain and Digital Humanities. He has published some 500 research articles and books and has got several international and national awards. https://seco.cs.aalto.fi/u/eahyvone/

Adeline Joffres is a research engineer at CNRS, Head of International Cooperation within the French Research Infrastructure Huma-Num. In particular, she works on the development and internationalisation of the Huma-Num's consortia. She is involved in the French coordination of the European research infrastructure DARIAH as well as the coordination of H2020 projects such as EOSC-Pillar (or previously PARTHENOS). Involved in the Coordinating Committee of the francophone association of Digital Humanities “Humanistica”, she is also working on the development of cooperation links between IR*Huma-Num and French higher education and research institutions abroad. https://www.huma-num.fr/adeline-joffres/

Peter Plassmeyer, Dr. phil., is an art historian and Director of the Mathematisch-Physikalischer Salon of the Dresden State Art Collections. He researches history of architecture, history of scientific and horological instruments and curates exhibitions. He introduced digital communication and educational tools in museums since 1996. Https://mathematisch-physikalischer-salon.skd.museum

Martijn Pronk is Head of Digital in the Van Gogh Museum, Amsterdam and former Head of Digital and Publishing in the Rijksmuseum, Amsterdam. His additional functions include seats on the Allard Pierson Foundation Board, the National Library of Israel Digital Strategy Advisory Committee and the WeAreMuseums Advisory Board. www.linkedin.com/in/phmpronk

Roberto Scopigno is a computer scientist, since 2019 appointed Director of CNR-ISTI. His research themes are 3D graphics and visual technologies, with main application to Cultural Heritage. He published more than 250 papers in international journals or conferences (Google Scholar h-index 60). He participated in many EC projects on ICT technologies and served as Editor in Chief for ACM J. on Computing and Cultural Heritage and Computer Graphics Forum. He has been the recipient of the

"Distinguished Career Award" issued by Eurographics Association in 2014. http://vcg.isti.cnr.it/~scopigno/

Gábor Sonkoly (Ph.D. EHESS, Paris; HDR, ELTE; Doctor of the Hungarian Academy of Sciences) is a professor of History and Head of the History PhD School at ELTE University, Budapest. He published 10 books and 80+ articles and book chapters on urban history, urban heritage and critical history of Cultural Heritage. He presented at 110+ international colloquia and was a guest professor in 14 countries. He is the academic coordinator of *TEMA+ European Territories: Heritage and Development* EMJMD, and Chair of the Panel for European Heritage Label. https://elte.academia.edu/G%C3%A1borSonkoly

APPENDIX

As a follow up of the Stakeholder meeting on 12 November 2021, an online survey was shared with thirty stakeholders representing the participating museums and cultural institutions to gather more input about the state and the process of digitization in their institutions. The survey was open from 20 January to 10 February 2022**. The summary of the answers is the following:**

- It is hardly surprising that every responding institution reported ongoing digitization activity. In addition, the experts who filled in the questionnaire were key researchers in the field of digital Cultural Heritage, meaning that institutions at the forefront of digitisation were predominant among the respondents. In most of the participating institutions, the digitisation rate of collections is between 30 and 50 percent, which is a very high rate.
- It is also not particularly surprising that in the digitised part of the collections, two-dimensional images, audio and video, and written documents are the most common media, but there were also examples of three-dimensional and even born digital material - albeit at a much lower rate.
- Institutions use a roughly equal share of external companies for the digitisation process and internal human resources.
- Institutions that are tackling the problem of physical and digital obsolescence are either developing solutions with the help of researchers or by the development of a digital preservation plan. Institutions are not uniform in how they collaborate on their digitised collections: although there are international project-based collaborations, it is just as common for these collections to be used only by individual researchers. The picture is much clearer when we ask what is the primary reason for publishing digital data: to make it available to the general public.
- Although the above objective is quite clear, the way in which data is published is very mixed, with all options from closed storage to open data being prominent.
- The responses indicate that the institutions that are more involved in digitisation do have a data policy, although only some of them have a data policy that includes open publication of data. This is all the more surprising as most institutions derive a low percentage of their revenues from the sale of digital materials. A significant proportion of institutions use a data repository to manage their data, and most of these are open-source products, but surprisingly there is a significant use of non-standard, in-house metadata sets. Not only digital objects but also the metadata used to describe them are not fully accessible to the public, with the majority of the institutions only publishing about half of the descriptive data freely. As far as the paradata sets are concerned, institutions are divided: about half of the institutions do not store them when digitising.
- It is encouraging that data enrichment is a priority for the majority of institutions, to which they are or would be devoting resources, and many of them are also involved in aggregation projects, but this is obviously also due to the high level of digital expertise of the stakeholders interviewed.
- The dominance of researchers in the reuse of data is very high, not only in relation to market or administrative uses, but even in relation to educational uses.