Deliverable D1.4 Data Management Plan 1

Deliverable type	DMP – Data Management Plan
Dissemination level	PU - Public
Due date (month)	M3
Delivery submission date	30 April 2023
Work package number	WP1
Lead beneficiary	Bilkent Universitesi Vakıf (BU)

Document Information

Project number	101094428	Acronym	CULTURATI
Project name	Customized Games and Routes for Cultural Heritage and Arts		
Call	HORIZON-CL2-2022-HERITAGE-01		
Topic	HORIZON-CL2-2022-HERITAGE-01-02		
Type of action	HORIZON-RIA		
Project starting date	1 February 2023	Project duration	36 months
Project URL	http://www.culturati.eu		
Document URL	https://culturati.eu/deliverables/		

Deliverable number	D1.4		
Deliverable name	Data Management Plan 1		
Work package number	WP1		
Work package name	Project Management and Coordination		
Date of delivery	Contractual	M3	Actual	M3
Version	Version 1.0		
Lead beneficiary	Bilkent Universitesi Vakif		
Responsible author(s)	Serdar Vural Öktem svoktem@gmail.com		
Reviewer(s)	Eda Gürel eda@tourism.bilkent.edu.tr Arzu Sibel İkinci aikinci@bilkent.edu.tr Özer Aydemir ozer@iotiq.de Metin Tekkalmaz metin@iotiq.de Angel Lagares alagares@nimbeo.com		

Short Description	This deliverable presents CULTURATI's first version of the Data Management Plan (DMP). It identifies data that will be handled throughout the project's lifecycle and how it will be generated, stored and backed up. It also addresses privacy, ethical, and security issues related to data management. This includes, among others, who owns data and is responsible for its preservation and which data will be shared and preserved for the future. This deliverable is prepared based on the

	Guidelines on FAIR Data Management in Horizon 2020 (European Commission, 2016).

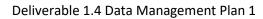| History of Changes			
Date	Version	Author	Remarks
20 April 2023	Draft 0.1	Serdar Vural Öktem	First version
30 April 2023	Final 1.0	Serdar Vural Öktem	Revised after review

Executive Summary

This deliverable presents CULTURATI's first version of the Data Management Plan (DMP). Thus, it deals with identifying data that will be handled throughout the project's lifecycle and how it will be created, stored and backed up. It also addresses privacy, ethical, and security issues related to data management. This includes, among others, who owns data and is responsible for its preservation and which data will be shared and preserved for the future.

The structure of this deliverable is based on the "Guidelines on FAIR Data Management in Horizon 2020" published by the European Commission (EC) in July 2016 (European Commission, 2023a). The content of this document will continuously be updated to describe new datasets (if any), refine the strategy, and include additional considerations before the final version is produced in M25.

Table of Contents

List of Tables

1. Introduction

The CULTURATI project is a Research Innovation and Action (RIA) project. It aims to develop, test, and validate an innovative online platform utilizing state-of-the-art digital and cutting-edge technologies to create content collectively for cultural heritage and arts across Europe. CULTURATI is a global content platform that delivers the content in its database to end-users (consumers of cultural heritage and arts) in customized games (Q&A games/treasure hunt-like game) and routes to attract and engage them with cultural heritage and arts more closely. It can be used by venue-based Cultural and Creative Industries (CCIs), professionals, and citizens. Thus, CULTURATI aims to close the gap between the demand and supply sides of the cultural and creative sectors by directly connecting them.

CULTURATI brings four aspects together: participative research with artifacts and information from the partners, and crowd data as part of its algorithm to be used to continuously train AI to provide the best options to the users. Consequently, data generation, collection, reuse, and preservation are crucial for CULTURATI's smooth running. It will need the collaboration of all the partners within the consortium: Bilkent University (BU), Haaga-Helia of Applied Sciences (HHU), Porvoon Kapunki (HHU-PK), Rahmi M. Koc Museum (RMK), Serdar Vural Oktem (SVO), Universita di Foggia (UNIFG), IOTIQ Gmbh (IOTQ), Nimbeo Estrategia e Innovacion S.L (NIMBEO), Universidad Carlos III De Madrid (UC3M), Ankara Governorship (GA) Meridaunia Soc. Cons. Arl (MSCA) Touch TD Ltd (TTD), Oxford Brooks University (OBU), Bleinheim Palace Heritage Foundation (BP). CULTURATI will follow the good scientific practice in preserving, managing, storing, documenting, and processing data adequately and ethically. Therefore, all consortium members should comply with this Data Management Plan in which specific procedures concerning data management are described.

This document, the first version of the CULTURATI Data Management Plan, describes the research data generated and collected during the project and the data (i.e. CCIs, the pilot sites digital assets, crowd data) that will be used throughout the project. The strategy to make data FAIR (Findable, Accessible, Interoperable, and Reusable) considers the balance between openness and protection of sensitive data.

In the future, updates and more details will be added as CULTURATI progresses to provide the guidelines and strategies to follow beyond the completion for exploiting data collected and generated within CULTURATI. A final version of this document will be released in M25 of the project.

2. Data Summary

2.1 Purpose

Data collection and generation are crucial for the following objectives. First, we will collect and generate data to enrich visitors' (consumers of cultural heritage and arts) experiences at pilot sites and validate the proposed solution in operational environments (indoors and outdoors).

By identifying visitors' preferences, we will provide them with rich experiences during their visit. Based on our conceptual model, visitors may be divided into two categories as omnivores and univores. In our online application, visitors may choose one or a few paths or gamification (Q&A Games/ Treasure Hunt Games). This data will help identify the characteristics of both visitor types.

Collecting crowd data to preserve the heritage and draw a less crowded path for visitors by employing CULTURATI's algorithms and AI will also help us enrich visitors' experiences. We will also collect data from the visitors to ensure they are satisfied with the proposed solution.

Second, CULTURATI is also a global content platform for anyone in public to co-create content involving cultural heritage and arts. Therefore, the system of CULTURATI will store the content (audio, video files etc) created collectively by the citizens.

Third, as a RIA project, the project will collect data from the users (content creators and end-users of cultural heritage and arts) to validate CULTURATI.

2.2 Types and Formats

CULTURATI aims to use a variety of data types and formats for research and technical aspects.

- **Rich media**: Text, image, audio, and/or video files about cultural heritage and arts coded under categories created by the content creators of pilot sites that can be shown on the web and mobile pages.
- **Metadata:** Data associated with describing artifacts in more detail, including their stories and locations. Hints for gaming will also be included in the metadata.
- **Pilot site maps/blueprint:** To provide optimized routes to the visitors, maps will be used. Maps will provide direction (Visual or Voice). To mark subareas such as rooms, exhibition areas, and shops, polygons will be used by placing points with coordinates on the vertices of the polygon. To determine the size of a polygonal area the size of at least one edge of the map

is needed. Similarly, the location of various items such as artworks and artifacts will be marked by collecting point coordinates.

- **User data:** There will be different types of users of CULTURATI; content creators, administrators of pilot sites, and end users (visitors). We will assign visitors a randomly generated unique ID to protect their identity. Content creators and administrators will log in to the system with a username and a password. In addition, their email addresses will be stored to protect their accounts.

- **Sensor data:** Wherever possible, sensors will be used to gauge crowd. When it is impossible, camera data will be collected in video or picture format. This data will be processed on our servers for privacy.

- **Research data:** Research data will be collected as part of our formative and summative evaluations. Data will be collected through online questionnaires and face-to-face interviews with the users of CULTURATI. We will collect questionnaire data through an online form at the cloud of CULTURATI based in Germany. In case of interviews, data will be entered manually.

- **Configuration data**: Configuration data is about system settings. This data will include sensitivities, our definitions for the crowd at locations in each pilot site for capacity management. Thus, it will consist of thresholds and data range boundaries.

Table 1: Types and formats of data generated/collected

	Type of Data	Description	Formats
1	Rich media	Text, Image, High-Resolution Photo, Audio and Video	JPG, TIFF, PNG TIFF, PNG MP3, MP4, 3D, txt, etc
2	Metadata	Subtitles, captions, categories, locations, tags, metadata, alt text	SRT, XML JSON, txt
3	Pilot site maps/blueprints	Sketch, map A graph defining vertexes and edges of the region and item locations with their dimensions	Spatial
4	User data	Credentials Preferences	Json txt, XML hashed/Encrypted text
5	Sensor data	Sensor/Camera Data	Json, XML, JPG, TIFF, PNG TIFF, PNG MP3, MP4
6	Research data	Information sheet Informed consent Feedback	DOC, DOCX, PDF XLS, WEB FORMS, SAV
7	Configuration data	System Setting	Txt, Json, XML

2.3 Origin

Pilot sites - Consortium partners: HHU-PK, MSCA, RMK, BP and AG are partners which will provide multimedia data, metadata, and maps. Also, sensor data will be collected at these sites depending on the technologies that will be chosen for CULTURATI. Research data will also be collected at these sites from the users of CULTURATI.

We will integrate the system of CULTURATI into the existing systems of the pilot sites, whenever possible. Otherwise, we will define import/export mechanisms to gather the data at the pilot sites.

User data: Users of CULTURATI include content creators, administrators at pilot sites, and end users of cultural heritage and arts. Administrators and content creators will be required to create an account with a username, a password, and an email address to use the system. On the other hand, end-users of CULTURATI will provide preferences to customize their experience. However, the system is going to assign randomly generated unique ids to protect privacy. End-users or visitors can create an account with an email address if they like the system to recognize them at subsequent visits. This can be on the Web or mobile platform.

2.4 Reuse

Most of the data will come from the multimedia, metadata, and maps from pilot sites and users of CULTURATI. This data can be stored by CULTURATI, or Application Programming Interfaces API's can be used to get the data from the provider. On the other hand, sensor/content data will be kept and reused for Machine Learning (ML). Since more data is better for AI decisions, and storage in the cloud is getting cheaper, they may be kept and used for a long time.

2.5 Expected Size

Many factors should be considered to estimate the size of data. For example for ML, normalization of the data should be considered. Moreover, using the the format for storing data is another consideration. Keeping data in ASCII format or in the database will require different sizes. Once alpha tests are completed, we will have a better estimation. Size depends on many factors, including the pilot site's size and how the content is created. For multimedia data especially, it depends on the method to be chosen. Depending on the size of the pilot area, the number of artifacts, and the resolution of the multimedia, we will estimate the size of the data.

During the project lifecycle, to validate CULTURATI, 10000 content (including text, audio, video, 3D files) will be created by five pilot sites across Europe (KPI). In addition, we will collect 15600 questionnaires and conduct 780 interviews with the users of CULTURATI to improve and validate it (KPI). The data will remain for eight years (five years after the project) in a cloud located in Germany to make it available to Cultural and Creative Industries (CCIs) and citizens in Europe and around the world.

2.6 Utility

Data will be useful for various target groups, including pilot sites, researchers, and the general public.

General Public: With CULTURATI, anyone who likes to create content for cultural heritage and arts will have a platform. Consequently, visitors and customers of cultural heritage and arts will receive rich information collectively co-created by anyone in public. Visitors can also avoid the crowd and visit the places they want by receiving customized information depending on their preferences. Internet of Technologies (IoT), specifically sensors, will be used to optimize the visitors' routes. With gamification, visitors can also enjoy their visit.

Pilot Sites: Pilot sites will have a better chance to know more about their visitors and preferences. They can also identify more or less popular artifacts and places on site by using crowd data. Therefore, they can design more creative marketing strategies and more satisfied visitors.

Researchers: Scientists will benefit from the feedback from participants, visitor behaviors, and their preferences. Gamification and AI can also provide rich research output.

3. Fair Management of Data

The consortium is committed to managing the digital and research data generated during the project responsibly, in line with the FAIR (Findable, Accessible, Interoperable, Reusable) principles bearing in mind that RDM (Research Data Management) is an essential element in any project that generates, collects or re-uses data to ensure that researchers can find, access and re-use each other's data, maximising the effectiveness and reproducibility of the research undertaken. Therefore, this section covers the four critical aspects of the DMP; findable, accessible, interoperable, and finally, reusable data.

3.1 Making Data/Research Outputs Findable

Findability of data and research outputs refers to the ability of others to locate and access our data and research products. It's an important aspect of data management planning because if our data and research products are not easily discoverable, it can limit the impact and reuse of our research outputs. To make data findable, we will pay attention to the following;

Use persistent identifiers (assign a Unique Identifier - PID): A Persistent Identifier is a long-lasting reference to uniquely identify a digital object, such as a dataset, publication, or software. We will use Persistent identifiers to provide a stable and reliable link to the object, even if it's location or metadata changes over time. There are several types of PIDs. We will use Digital Object Identifiers (DOIs) to provide a long-lasting and actionable reference to digital objects, making them easier to find and cite.

Provide descriptive and standardized metadata: The data will be enriched with metadata so it can be easily discovered. Metadata refers to descriptive information or data that provides context, structure, and meaning to other data or information. Thus, we will describe our digital resources' characteristics, attributes, and relationships. Consequently, it will help others find and understand our data. Metadata may include the title, author, description, date, and keywords. Data findability will also be facilitated by the Data Lake (WP2 – Task 4).

Use standardized file formats and naming conventions: We will use standardized file formats outlined in Table 1 to ensure others can easily find and use our data. In addition, naming conventions are also useful for this purpose. Accordingly, we will follow the Institute of Electrical and Electronics Engineers (IEEE) standards. For example, the IEEE 2791-2020 Standard for Data Documentation in Health Research provides guidance on the content and format of metadata and other documentation to be included in datasets for health research. Similarly, the IEEE 2600 Standard for Enabling Privacy in Data Management Systems provides a framework for managing privacy risks associated with collecting, storing, and sharing data (IEEE, 2023). We will follow these standards as guidelines to ensure findability, privacy, and compliance with relevant laws and regulations.

Deposit our data in a repository: We will deposit our digital and research data on APERTA, the Scientific and Technological Research Council of Türkiye's open-access repository for researchers. The data on APERTA is identified by a persistent identifier and a stable resolvable link to where data sets can be directly accessed. All deliverables of our CULTURATI project open to the public will be available on APERTA. In addition, our data and system files will also be stored in a cloud located in Germany.

To ensure findability, documents will also be versioned. Data and metadata can be stored in NoSQL or relational databases, and both have Version Control Systems (VCS). For this purpose, there will be keyword-based and full-text search functions. Search can be done online and from mobile devices.

3.2 Making Data Openly Accessible

As an integral part of our methodology, we will implement open science practices in various ways to make our research outputs transparent, available, accessible, broad, and sustainable. To this end, we will pay attention to the following;

Use public-open access platforms and choose an appropriate access level: The primary objective of the CULTURATI project is to create a global content platform, thus making all information about cultural heritage and arts open to users in various formats, including pictures, audio and video files, stories about the artifacts and places. This information will be online and can be accessed from the internet.

Research data will be available for researchers. The consortium will benefit from Open Research Europe to make our findings openly accessible to a broader community. In line with open science practices, most of the deliverables of CULTURATI are also open to the researchers and the public through APERTA and the website of CULTURATI culturati.eu (see section 5.4 Deliverables in the D1.1 Project Handbook). By default, material deposited in APERTA is openly accessible worldwide over the Web. However, open access to the content of the repositories will be subjected to IPRs. Therefore, each partner will carefully assess on a case-by-case basis which results – and to what extent – can be made public following the Consortium Agreement (CA).

In addition, for privacy, user credentials will not be open to anyone, even to the developers of CULTURATI. They will be stored encrypted during the login. When the user logs into CULTURATI, it will be encrypted and compared with encrypted data from the database. On the other hand, user preferences will be open to only consortium members. Although the backend software will use this data to draw routes and curate customized content, anonym data will be available for researchers and the administrators of the pilot sites.

Use appropriate licensing: Making data accessible and public does not mean uncontrolled distribution of the data used or generated in the project. Data licenses help to protect data by one or more IP

rights and address different aspects relevant to data handling such as delivery, security, maintenance, control, usage conditions, etc. There are many different open data licenses and the most common ones include Open Data Commons Public Domain Dedication and License (PDDL), Creative Commons Attribution 4.0 International (CC-BY), Community Data License Agreement – CDLA Permissive-2.0, Open Data Commons Attribution License (ODC-BY), Creative Commons Attribution-ShareAlike 4.0 International (CC-BY-SA), Community Data License Agreement – CDLA-Sharing-1.0, Open Data Commons Open Database License (ODC-ODbL). In the CULTURATI project, a proper license will be selected for each dataset to be openly accessible, and the corresponding dataset will be made available under that license.

Publish in open-access journals: Publishing in open-access journals increases the visibility of our research and can also lead to improved data reuse. Therefore, we will publish in open-access journals. Our budget includes the cost of it.

Provide clear instructions for access: We provide access to our data through CULTURATI's website (culturati.eu) located at Bilkent Universitesi Vakif. We provide direct links to the data on APERTA through the project website.

Provide the data in a usable format: Our data is provided in a format others can easily use defined in Table 1.

Provide accessibility for users with disabilities: The data will be accessible to users with disabilities. This will include providing alternative formats, such as audio, large print, or video captions.

3.3 Making Data Interoperable

Data interoperability refers to the ability of different systems and applications to exchange and use data in a consistent and standardized way. It ensures that data can be shared and reused across different platforms, applications, and organizations without losing meaning or integrity. Interoperable data is typically formatted using open, standardized, and machine-readable formats and protocols, such as CSV, JSON, XML, or RDF. This allows data to be easily processed, analyzed, and combined with other data, regardless of the system or application that produced it.

However, achieving data interoperability can be challenging, as it requires aligning different data models, standards, and vocabularies, as well as addressing issues related to data quality, security, and

privacy. To ensure interoperability, it's important to establish clear data governance policies, metadata standards, and data management practices and to adopt open and widely accepted data formats and protocols. To this end, we will pay attention to the following;

Using standard data formats, vocabularies and ontologies: We will use standard data formats that are widely accepted and easily recognized, as mentioned in Table 1. This will ensure that others can easily exchange and use our data. We will also use standard vocabularies and ontologies to ensure that our data use consistent terminology and is easily understood by others.

To this end, in particular, for the museums among the pilot sites of CULTURATI, we will use the Conceptual Reference Model (CRM) developed by the International Committee of Documentation (CIDOC) of the International Council of Museums (ICOM) is a well-known and extensively used semantic model based on earlier standards. CIDOC-CRM establishes relationships among implicit and explicit concepts for cultural heritage documentation to transform isolated and inhomogeneous metadata into a valuable and coherent global resource (CIDOC-CRM, 2023). We will also refer to the Europeana Data Model (EDM) by the Europeana network (Europeana, 2023). It aims to guarantee the preservation of the original data from the diverse metadata schemas while accommodating the wide variety of community standards for museums, archives, and digital libraries.

Using unique identifiers: We will use unique identifiers for our data, such as Digital Object Identifiers (DOIs), which allow others to reference and cite our data easily (see section 3.1 above).

Following data documentation standards: We will document our data using metadata standards and provide detailed descriptions, including information about how it is collected and processed (see section 3.1 above and 7.2 Documentation in the D1.1 Project Handbook).

Providing documentation and code: Provide documentation and code related to our data, such as codebooks or data dictionaries, to make it easier for others to understand and use our data. For this purpose, a data dictionary will be created; it will serve as a reference and description of each data element. The dictionary will be based on the data model; it will be based on the detailed definition and documentation of the data model as mentioned in the Grant Agreement.

Using open APIs: We will use open Application Programming Interfaces (APIs) to facilitate access to our data and enable integration with other data sources.

3.4 Increase Data Reuse

Data usability refers to the ease with which users can understand, access, process, and analyze data. Usable data is typically well-structured, documented, and formatted, with clear and concise metadata, data dictionaries, and other supporting materials.

The CULTURATI consortium is committed to exploiting and sharing the outcomes generated (knowledge, technologies, data, etc.) in the framework of the activities defined in the work plan. To this end, free access, public deliverables, open publications, reports, or access to research data indicate this commitment. However, all the aspects related to IPR should be thoroughly assessed according to the Grant Agreement and the Consortium Agreement.

The cloud platform and web applications will be kept updated and running for at least five years after the completion of the project, allowing the consortium to improve the functionalities, add new content and initiate the commercial exploitation. During this period, the access to the repositories will be supported as usual, so that third parties will be able to reuse data.

To increase data reuse, we will pay attention to making our data openly available, providing clear and detailed documentation, and using standard file formats and data structures as described in the sections above.

4. Allocation of Resources

The estimated costs for making data FAIR were already considered when drafting the budget for the project before being submitted to the European Commission. Implementing dedicated spaces in data repositories was deemed an essential part of the approach and the access to the collected and generated data. Consequently, the consortium does not need to distinguish or consider additional resources to fulfill the expectations.

We assessed our data and identified the resources for the project preparations. All estimated costs, including hardware and software, data storage and management, and personnel, were in the project budget. In particular, the budget includes the cost of the cloud located in Germany for eight years (including five years after the project) and open-access journals. Necessary resources have been allocated to each prioritized activity, adjusting timelines, budget allocations, and staffing. We will monitor the progress regularly to ensure that activities are completed on time and within budget. If necessary, we adjust the allocation of resources as needed.

The DMP is a live document and will be updated based on the project progress and needs. Among the partners, Serdar Vural Öktem is involved in the project for data management and protection due to his expertise.

5. Data Security and Protection

While data security concerns protecting the data itself, data protection is about protecting the privacy rights of individuals whose data is being processed. Both are important for ensuring compliance with relevant laws and regulations, such as the General Data Protection Regulation (GDPR).

To ensure data security in the CULTURATI project, we will implement various technical and organizational measures, such as access controls, encryption, firewalls, intrusion detection and prevention systems, backup and recovery procedures. To this end, we will do the following;

- We will use a cloud server located in Germany and will be secured by several firewalls,
- To comply with good scientific conduct, the primary data within CULTURATI will be stored for eight years; thus the data will be stored five years after the end of the project,
- We will take complete and redundant backups every day. In addition, an older version will be saved when a modification is done. In case of accidental deletion or modification, the data from the most recent backup will be automatically restored,
- Only administrators will have the right to delete or modify the information included in the datasets. In such case, the user responsible for its creation will be notified before doing so,
- All personal user data (we will collect only usernames, passwords, and email addresses) will be protected by a two-factor authentication in the cloud with access only by authorized team members,
- System users' and participants' usernames, passwords, and email addresses will not be open-access and shared by third individuals,
- For data protection, research data and system data will be stored on different servers in the cloud in the EU,
- There will be encrypted data storage in which data will be stored according to necessity,
- For information security, we will comply with ISO/IEC 27001, a widely recognized international standard that sets out the requirements for implementing, maintaining, and continually improving an information security management system (ISMS) (ISO, 2023).

On the other hand, to ensure data protection, we will safeguard personal data from unauthorized or unlawful processing, disclosure, or access. To this end, we will implement policies, procedures, and technologies to ensure that data privacy rights are respected and that data subjects are informed and in control of their data. For this purpose, we will pay attention to the following;

- For data protection, it's important to follow data minimization principles, purpose limitation, transparency, accuracy, security, and accountability. This includes only collecting and processing data that is necessary and proportionate for the intended purpose, informing data subjects about the processing of their data, ensuring the accuracy and completeness of data, and protecting data from unauthorized access, use, or disclosure. Accordingly, we will only collect data from the users of CULTURATI (content creators, online and on-site end-users, and managers of pilot sites) voluntarily and anonymously using randomly generated unique ids. Thus, we will pseudonymize data. Pseudonymization replaces identifiable information, such as a name or an email address, with a pseudonym, such as a randomly generated identifier.This code will be attached to the digital records of all users of CULTURATI, including the participants, as part of our research data. We intend to keep all our research data anonymous to prevent potential mis-use. For more details, refer to the Deliverable D1.3 Ethics Assessment Report.

- Data protection also involves outlining the procedures for preserving the data for the long term, including how data will be archived, the format in which it will be archived, and any metadata required for long-term preservation. Thus in CULTURATI, our data formatting and metadata provision will conform to the" guidelines on data protection for archive services" of the European Archives Group (European Commission, 2023b).

- As part of our Deliverable D2.1 System Design and Specification, the Appendix will include the Security Architecture Design consisting of the following;
  - Tools for authentication, authorization, privacy, and integrity,
  - Network security protocols, with their sample configuration,
  - Encryption methods that will be used,
  - A security guide for system admins.

Finally, for data security and protection, all data of CULTURATI will be kept in compliance with EU's General Data Protection (GDPR) approved by the EU Parliament on 14 April 2016, the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018, and the Law on the Protection of Personal Data" No. 6698 came into force 7 April 2016 in Türkiye.

6. Ethical Aspects

The Deliverable D1.3 Ethics Assessment Report outlines guidelines and procedures for addressing various ethical issues that may arise during our project, particularly regarding research practices and technology development. As set out in the European Code of Conduct for Research Integrity of ALLEA (All European Academies, 2017), the fundamental principles of research integrity of honesty, reliability, respect, and accountability will guide our approach to these issues.

As discussed in more detail in the Deliverable D1.3 Ethics Assessment Report, we will prioritize **transparency and disclosure** to promote **honesty** throughout our research process. This includes our various open science practices and data management procedures. In addition, we will adhere to rigorous research standards to ensure **reliability**, including robust methodology, data collection and analysis procedures, careful documentation, and peer review. We will also take steps to mitigate potential sources of error, such as sampling bias. We also address obtaining ethical approvals. Moreover, to demonstrate **respect**, we will prioritize the welfare and dignity of all individuals involved in our research, including study participants, research staff, and other stakeholders. This includes obtaining informed consent, protecting privacy and confidentiality, and avoiding harm or exploitation. Furthermore, to promote **accountability**, we will establish clear roles and responsibilities, set expectations for ethical behavior, and establish mechanisms for oversight and review. We will also ensure that our research adheres to applicable **laws, regulations, and professional codes of conduct**.

By following these principles, we will ensure that our project is conducted ethically and with the highest level of integrity.

Conclusion

This deliverable describes the initial strategy to deal with data collection, generation, management and exploitation. This is a living document and will be updated continuously. This will allow us to improve and fine tune the preliminary actions (including the identification of datasets) as well as implement new measures if necessary. The final version will be delivered in M25.

The section dealing with the data summary presented the purpose, types and formats, reuse, origin, expected size, and utility of the data collected and generated in CULTURATI. Various technical and organizational measures were defined based on the diverse characteristics and needs of the project.

Our Data Management Plan is prepared based on the FAIR principles, essential components of good research practice. By making data Findable, Accessible, Interoperable, and Reusable, we can ensure that our data is high-quality, well-managed, secure, and available for reuse by others. In addition, the Data Management Plan is also based on the European Code of Conduct for Research Integrity of ALLEA (All European Academies, 2017); the fundamental principles of research integrity of honesty, reliability, respect, and accountability.

In addition to these principles, ensuring the security and protection of research data is also a critical aspect of our data management plan. Our data security measures include access control, encryption, and regular backups are essential to protect data against unauthorized access, loss, or damage. Similarly, we covered data protection measures, including anonymization, pseudonymization, and informed consent can help to ensure that personal data is handled appropriately and in compliance with legal and ethical requirements.

References

ALLEA (2017, April 17). The European Code of Conduct for Research Integrity Revised Edition. https://allea.org/code-of-conduct/

CIDOC-CRM (2023, April 28). What is the CIDOC CRM? http://www.cidoc-crm.org/

European Commission (2023a, April 17). Horizon 2020 Online Manual – Data Management. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

European Commission (2023b, April 19). European Archives Group. https://commission.europa.eu/about-european-commission/service-standards-and-principles/transparency/access-documents/information-and-document-management/archival-policy/european-archives-group_en

Europeana (2023, April 23). Discover Europe's digital cultural heritage. http://www.europeana.eu

IEEE (2023, May 2). Raising the World's Standard. https://standards.ieee.org/

ISO (2023, May 5). ISO/IEC 27001 Information security management systems. https://www.iso.org/standard/27001