



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870811



D1.2 Data Management Plan

Deliverable information	
WP	WP1
Document dissemination level	PU Public
Deliverable type	ORDP
Lead beneficiary	UNIBO
Contributors	ALL
Date	10/05/2022
Document status	Final version
Document version	V1.1

Disclaimer: The communication reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains

INTENTIONALLY BLANK PAGE

Project information

Project start date: 1st of May 2020

Project Duration: 36 months

Project website: <https://spice-h2020.eu>

Project contacts

Project Coordinator

Silvio Peroni

ALMA MATER
STUDIORUM -
UNIVERSITÀ DI
BOLOGNA

Department of Classical
Philology and Italian
Studies – FICLIT

E-mail:

silvio.peroni@unibo.it

Scientific Coordinator

Aldo Gangemi

Institute for Cognitive
Sciences and Technologies
of the Italian National
Research Council

E-mail:

aldo.gangemi@unibo.it

Project Manager

Adriana Dascultu

ALMA MATER
STUDIORUM -
UNIVERSITÀ DI
BOLOGNA

Executive Support Services

E-mail:

adriana.dascultu@unibo.it

SPICE consortium

No.	Short name	Institution name	Country
1	UNIBO	ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA	Italy
2	AALTO	AALTO KORKEAKOULUSAATIO SR	Finland
3	DMH	DESIGNMUSEON SAATIO - STIFTELSEN FOR DESIGNMUSEET SR	Finland
4	AAU	AALBORG UNIVERSITET	Denmark
5	OU	THE OPEN UNIVERSITY	United Kingdom
6	IMMA	IRISH MUSEUM OF MODERN ART COMPANY	Ireland
7	GVAM	GVAM GUIAS INTERACTIVAS SL	Spain
8	PG	PADAONE GAMES SL	Spain
9	UCM	UNIVERSIDAD COMPLUTENSE DE MADRID	Spain
10	UNITO	UNIVERSITA DEGLI STUDI DI TORINO	Italy
11	FTM	FONDAZIONE TORINO MUSEI	Italy
12	CELI	CELI SRL	Italy
13	UH	UNIVERSITY OF HAIFA	Israel
14	CNR	CONSIGLIO NAZIONALE DELLE RICERCHE	Italy

Executive summary

This deliverable introduces the Data Management Plan (DMP) of SPICE. The DMP outlines the strategy for the management and accessibility of data generated in the project, including software and data from experiments, surveys, and interviews.

Document History

Version	Release date	Summary of changes	Author(s) - Institution
V0.1	28/10/2020	First draft released	UNIBO
V0.2	30/10/2020	Revision by the partners and internal review	ALL
V0.3	03/11/2020	Final internal review	OU
V1.0	04/11/2020	Final version submitted	UNIBO
V1.1	10/05/2022	Revised version	UNIBO

Table of Contents

Project information	3
Project contacts.....	3
SPICE consortium	3
Executive summary	4
Document History	5
Introduction	7
Data summary.....	7
FAIR Data.....	17
FINDABILITY	17
ACCESSIBILITY	18
INTEROPERABILITY	19
REUSABILITY	21
Allocation of resources.....	21
Data security	22
Ethical aspects.....	22

Introduction

The goal of the Data Management Plan (DMP) is to ensure that results and scientific products resulting from the project are as open and accessible as possible. SPICE data and software will be available from the project web portal, dedicated infrastructures developed by SPICE consortium, and through selected dissemination infrastructures such as GitHub, ZENODO, and institutional repositories for data and publications.

Data collected, produced and managed during the project mainly cover aspects related to museum visitors' encounters with artworks, such as emotions, interpretations, and opinions. Data are generated and/or leveraged in engagement activities relevant to the five project pilots (WP7). This information is collected in workshops (WP2), engagement activities planned in the aforementioned five case studies (WP7) and mined from social media platforms (WP4). Existing and bespoke ontologies are developed so as to organize collected data and enable further analysis and reuse (WP6).

Software solutions developed by SPICE are primarily meant to allow stakeholders (e.g. copyright holders, museum curators, developers, end users) to manage and share collected data according to different levels of privacy and copyright restrictions (WP4). Secondly, bespoke software solutions are developed in order to apply reasoning methods on collected data (WP6), visualize and explore users' opinions (WP5) and derive information on citizens' communities and reflection for recommendation purposes (WP3).

Due to privacy concerns and commercial opportunities for SME involved in the project, not all the data and software solutions will be fully openly available, even if some of them could be made available to the consortium for the purpose of the research run in SPICE. In such cases, the software will be made available to the consortium as a service. Following the Software as a Service (SaaS) model, the software will expose REST APIs to other software clients developed by the consortium. As a result, the functionalities provided by the software are available to the consortium, but the source code of the software remains private. An example of this solution was adopted for the Semantic Annotator (cf. D3.2 and D3.4). Museum visitors' identities and user-generated data will be mostly anonymized so as to open as much data as possible without disclosing personal information and comply with ORDP requirements. Whenever applicable, users will be asked to give their consent in order to publish their identities. This will be done in full compliance with any European and national legislation relevant to the country where the data collections are taking place. Only data for which user's consent has been asked will be fully disclosed and publicly available for reuse along with anonymized data.

Data summary

The following table includes an overview of datasets and software solutions developed by SPICE project partners ordered by responsible WP.

Acronyms used: Type = S (Software) / D (Dataset); Status = A (Available) / O (Ongoing) / NA (Not yet available). The DMP identifier is used through the document to identify research outputs.

Type	DMP identifier	Short name	WP	WP Task	Status
D	Workshop	SPICE Participatory Workshop Database	WP2	1-2	O
S	Annotation	Semantic annotation service of social curatorial products	WP3	2	A
S	UM	User Model	WP3	1	O
S	Community	Community visualization tool	WP3	3	O
S	Recomm	Recommender	WP3	4	O
S	LDH	SPICE Linked Data Hub	WP4	1-4	A

D	LDHReg	SPICE Linked Data Hub Registry	WP4	1-4	O
D	IGmuseum	Instagram museums data	WP4	3	O
S	IGclassifier	Classifier of social media users' interactions	WP4	3	NA
S	IUcitizen	Interface components for citizen curation	WP5	1-3	O
S	Reasoner	On Demand Ontology Reasoner and Server	WP6	1	A
D	Ontologies	SPICE Ontologies network	WP6	3	A
D	IMMAdata	IMMA Pilot Study dataset	WP7	3	O
S	IMMAPilot	IMMA Pilot Study software	WP7	3	O
D	MuOpinion	Museum visitor opinion dataset	WP7	3	O
S	HaifaPilot	Haifa case study web application	WP7	3	O
D	Demog	End users' demographics, preferences and results from activity analysis	WP7	3	O
D	Ethnog	Autoethnographic and duoethnographic accounts of material culture	WP7	3	O
D	PortableMu	Surveys documenting responses to prototypes	WP7	3	O
S	VRpopup	VR experience of the Pop-up Museum	WP7	3	O
D	FTMdata	User-generated data on the web	WP7	3	O

Table 1 Overview of datasets and software produced in SPICE

In the following tables we further describe the outlined research outputs, including persistent identifiers, creators, contributors, rights holders, description and intended purpose, data “utility”, data types and formats.

Short name	<i>SPICE Participatory Workshop Database</i>
DMP identifier	Workshop
Identifier	-
Creators	<i>Luis Emilio Bruni, Aalborg University; Mark Maguire, Irish Museum of Modern Art Company</i>
Contributors	<i>Luis Emilio Bruni, Aalborg University; Luca Simeone, Aalborg University; Thomas Anthony Pedersen, Aalborg University; Diaz-Kommonen Lily, Aalto University; Vishwanath Gautam, Aalto University; Leena Svinhufvud, Design Museum Helsinki; Mark Maguire, Irish Museum of Modern Art Company; Anna Follo, Fondazione Torino Musei; Giorgia Rochas, Fondazione Torino Musei; Federica Sesia, Fondazione Torino Musei; Alan Wecker, University of Haifa and Hecht Museum; Tsvi Kuflik, University of Haifa and Hecht Museum; Joel Lanir, University of Haifa and Hecht Museum; Anna Maria Marras, University of Turin; Rossana Damiano, University of Turin; Cecilia Marchisio, University of Turin; Barbara Bruschi, University of Turin; Enrico Dolza, University of Turin; Antonio Pizzo, University of Turin; Stefano De Giorgis, University of Bologna; Francesca Tomasi, University of Bologna; Bruno Sartini, University of</i>

	<i>Bologna; Silvio Peroni, University of Bologna; Marilena Daquino, University of Bologna.</i>
Rights Holders	<i>Aalborg University; Aalto University; Design Museum Helsinki; Irish Museum of Modern Art Company; Fondazione Torino Musei; University of Haifa and Hecht Museum; University of Turin; University of Bologna</i>
Description and purpose	<p>Description: Survey results, interview transcripts, with general data about participants' personal views and perspectives related to themes and subjects identified in particular artworks proposed in cultural heritage activities.</p> <p>Purpose: Collect requirements, and assess success and challenges of the project's activities, that are shared with relevant WPs. Used anonymously externally to illustrate the development of the project.</p>
Data utility	Useful for other WPs to define requirements for data modelling, reasoning (WP6), visualization and analysis (WP3, WP5).
Data types and formats	Qualitative and quantitative, collected raw data. Formats: txt, docx, pdf, xlsx, jpg, png

Table 2 SPICE Workshop Database

Short name	<i>Semantic annotation of social curatorial products</i>
DMP identifier	Annotation
Identifier	https://sophiaanalytics.saas.celi.it/
Creators	<i>Alessio Bosca, CELI</i>
Contributors	<i>Alessio Bosca, CELI; Chiara Albano, CELI</i>
Rights Holders	<i>CELI</i>
Description and purpose	<p>Description: Annotation Service for multilingual user generated contents. It enriches multilingual textual contents from museum visitors with semantic annotations (Entities and Key Concepts, Emotions and Sentiment, Opinions). The supported Languages are English, Finnish, Hebrew, Italian, Spanish. The service output consists in a JSON LD document and it will be used by other WP3 tasks in the process of generating users and communities models.</p> <p>Purpose: Extract structured data from natural language texts generated by users.</p>
Data utility	Data extracted by means of the tool are mainly used in the context of WP3 to support the design of user and community models.

Table 3 Semantic annotation service of social curatorial products

Short name	<i>User Model</i>
DMP identifier	UM
Identifier	-
Creators	<i>Alan J. Wecker, University of Haifa</i>
Contributors	<i>Tsvi Kuflik, University of Haifa; Joel Lanir, University of Haifa; Iris Reinhalz-Berger, University of Haifa</i>
Rights Holders	<i>University of Haifa (UH)</i>
Description and purpose	Description: Techniques for identifying user groups and communities from the content provided by users. Content provided by museum visitors is analysed by

using unsupervised machine learning techniques for identifying key concepts representing individuals and groups. Multilingual text analytics will be applied to user content in order to perform named entity recognition, opinion mining and emotion recognition.

Purpose: To create community models based on content similarity.

Data utility Preliminary to other activities in WP3 (see [Recommender](#)) and meant to be used in WP7 case studies.

Table 4 User Model

Short name	<i>Community visualization tool</i>
DMP identifier	Community
Identifier	-
Creators	M Belén Díaz Agudo , Universidad Complutense de Madrid; Guillermo Jiménez Díaz , Universidad Complutense de Madrid
Contributors	-
Rights Holders	Universidad Complutense de Madrid
Description and purpose	Description: A interactive tool using clustering techniques for identifying commonalities and variabilities among user communities. The tool is an experimental environment where we can explore different techniques and different levels of abstraction. The tool will perform experimentation using synthetic, real user data and content from the different case studies about the artifacts, and real user's contributions. It will be evaluated with user studies, by providing a visual interface and interactive options. Purpose: To visualize communities and aggregation of users so as to understand patterns and differences.
Data utility	The tool is mainly used in the context of WP3 to support data exploration and design of recommendation models tailored for different communities.

Table 5 Community visualization tool

Short name	<i>Recommender</i>
DMP identifier	Recomm
Identifier	-
Creators	Alan J. Wecker , University of Haifa
Contributors	Tsvi Kuflik , University of Haifa; Joel Lanir , University of Haifa; Iris Reinhalz-Berger , University of Haifa
Rights Holders	University of Haifa
Description and purpose	Description: Recommending system based on the user and community models built from the analysis of textual contents (WP3) and terms from the concepts ontology (WP6) to help users to explore and discover different and even conflicting points of view. Purpose: To suggest content to be “consumed” by community members - content related to their communities (intra-community) or other communities (inter-community e.g. not only similar but also diverse and possibly conflicting points of views).
Data utility	To be used in case studies (WP7) for engaging with museum visitors.

Table 6 Recommender

Short name	<i>SPICE Linked Data Hub</i>
DMP identifier	LDH
Identifier	https://spice.kmi.open.ac.uk/
Creators	<i>Enrico Daga, The Open University</i>
Contributors	<i>Enrico Daga, Open University; Jason Carvalho, Open University; Paul Mulholland, Open University; Luigi Asprino, University of Bologna; Marilena Daquino, University of Bologna</i>
Rights Holders	<i>The Open University</i>
Description and purpose	Description: The SPICE LDH is a repository and mediator in the linked data ecosystem of the SPICE project. Purpose: Hub for accessing and managing data policies of all SPICE datasets. Its functionalities also include services related to the research output of WP4.
Data utility	LDH is a policy and privacy-aware environment supporting all SPICE partners producing, managing, and publishing data.

Table 7 SPICE Linked Data Hub

Short name	<i>SPICE Linked Data Hub Registry</i>
DMP identifier	LDHReg
Identifier	-
Creators	<i>Enrico Daga, The Open University</i>
Contributors	-
Rights Holders	<i>The SPICE Consortium</i>
Description and purpose	Description: The registry of the linked data resources produced in the SPICE project. Data are mainly of qualitative nature, including cataloguing information of the datasets generated by SPICE partners. Purpose: The dataset is generated to support the Data catalogue management system included in the SPICE Linked Data Hub.
Data utility	All the partners producing or consuming data in SPICE; Developers and stakeholders that want to access the data.
Data types and formats	Qualitative, newly generated raw data. Formats: json-ld

Table 8 SPICE Linked Data Hub Registry

Short name	<i>Instagram museums data</i>
DMP identifier	IGmuseum
Identifier	-
Creators	<i>Enrico Daga, The Open University</i>
Contributors	<i>Marilena Daquino, University of Bologna</i>
Rights Holders	<i>The Open University</i>
Description and purpose	Description: Collected data from users' posts on Instagram referencing artworks owned by museums. Information on users' sentiment, emotion, and intention are extracted, and posts classified accordingly. Data are reengineered in JSON-LD and enriched with stable links to museums' cataloguing records. The data are evaluated by evaluating the software (see IGclassifier) for producing the dataset. Purpose: Characterize users' behaviours on social media platforms with regard to their personal encounters with artworks. Use such information to develop new methods for discoverability across museum collections.

Data utility	Data can be reused by project partners and museums (e.g. social media campaigns, resources recommendation), and by developers in social semantic web applications.
Data types and formats	Quantitative and qualitative, derivative and raw, processed data collected from Instagram APIs. Formats: json-ld

Table 9 Instagram museums data

Short name	<i>Classifier of social media users' interactions</i>
DMP identifier	IGclassifier
Identifier	-
Creators	Marilena Daquino , University of Bologna
Contributors	Enrico Daga , The Open University
Rights Holders	University of Bologna
Description and purpose	Description: Software for image matching and text classification according to sentiment/emotion/intention. The software allows to produce enriched datasets including links between museum objects and users' posts on Instagram, emotions, and users' perspectives. User studies will be developed to evaluate precision of the classifier. Purpose: Foster discoverability across museum collections using user-generated data.
Data utility	The LDH Registry, to include meaningful links between assets; Museums, to include meaningful user-generated data; developers in social semantic web applications.

Table 10 Classifier of social media users' interactions

Short name	<i>Interface components for citizen curation</i>
DMP identifier	IUcitizen
Identifier	-
Creators	Pedro González , PadaOne Games
Contributors	Pedro Gonzalez (PG) , Marco Gómez (PG) , Guillermo Laseca (GVAM) , Jaime Solano (GVAM)
Rights Holders	PadaOne Games, GVAM
Description and purpose	Description: The interface components for citizen curation will consist of a number of independent interface modules for (1) different types of interpretative tasks, such as tagging, collecting, and storytelling, for specific communities, including those with disabilities; (2) for browsing and exploring across a set of interpretations to understand the range of responses. Purpose: Support users in developing and discovering different opinions.
Data utility	Components will be integrated into the final version of the use cases (WP7) in SPICE.

Table 11 Interface components for citizen curation

Short name	<i>On Demand Ontology Reasoner and Server</i>
DMP identifier	Reasoner
Identifier	-
Creators	Luigi Asprino , University of Bologna; Antonio Lieto , University of Turin

Contributors	-
Rights Holders	<i>University of Bologna, University of Turin</i>
Description and purpose	<p>Description: A service-oriented application that accept as input the IRI identifying an entity or a SPARQL query about that entity and returns a list of facts about the entity, either explicitly asserted or deduced using the SPICE ontologies network. This component reuses state-of-the-art reasoners like Pellet and HermiT, available through frameworks such as OWLAPI, OntoAPI, JENA and uses the Apache Jena Fuseki 2 server for exposing its reasoned models to other services. The ontology reasoner will also exploit and readapt the DENOTER system to recommend novel items/artwork to the users based on the interpretation/reflections loops. The software will be evaluated through a proof-of-concept deployed in the context of the case studies.</p> <p>Purpose: Infer knowledge about entities stored in the knowledge base of SPICE.</p>
Data utility	Complements activities of WP3 in data sense-making.

Table 12 On Demand Ontology Reasoner and Server

Short name	<i>SPICE Ontologies network</i>
DMP identifier	Ontologies
Identifier	-
Creators	<i>Aldo Gangemi, Italian National Research Centre (CNR); Luigi Asprino, University of Bologna; Marilena Daquino, University of Bologna; Stefano de Giorgis, University of Bologna; Silvio Peroni, University of Bologna; Bruno Sartini, University of Bologna</i>
Contributors	-
Rights Holders	<i>University of Bologna, Italian National Research Centre (CNR)</i>
Description and purpose	<p>Description: The ontology consists of two modules: 1) the citizen curation ontology that describes the process by which citizens will produce and share interpretations of museum objects through activities conducted as part of the social curation scripts. 2) The interpretation ontology describes the interpretations produced by the citizen curation activities. The modules will be designed by using the eXtreme Design methodology.</p> <p>Purpose: The ontologies network is aimed at connecting the representation of museum objects with interpretations provided by citizens and curators through curation activities.</p>
Data utility	All WPs producing data.
Data types	Qualitative generated data.
Data types and formats	Formats: n-triple, turtle, rdf/xml, json-ld

Table 13 SPICE Ontologies network

Short name	<i>IMMA Pilot Study dataset</i>
DMP identifier	IMMAdata
Identifier	-
Creators	<i>Paul Mulholland, The Open University</i>
Contributors	<i>Enrico Daga, The Open University; Mark Maguire, Irish Museum of Modern Art Company</i>
Rights Holders	<i>The Open University</i>

Description and purpose	Description: The dataset includes survey results, interview transcripts, data on users’ activities, and curatorial contributions. The dataset is evaluated according to methods designed in WP2.
Data utility	Purpose: Evaluate results of Irish pilot in WP7. Pilot in WP7.
Data types and formats	Qualitative and quantitative, raw and derivative, processed data. Formats: json-ld

Table 14 IMMA Pilot Study dataset

Short name	<i>IMMA Pilot Study software</i>
DMP identifier	IMMAPilot
Identifier	-
Creators	<i>Paul Mulholland</i> , <i>The Open University</i>
Contributors	<i>Enrico Daga</i> , <i>The Open University</i>
Rights Holders	<i>The Open University</i>
Description and purpose	Description: the software for manipulating and reengineering data We follow the evaluation methodology as designed by WP7 (T7.2). Purpose: Evaluate results of Irish pilot in WP7.
Data utility	Pilot in WP7.

Table 15 IMMA Pilot Study software

Short name	<i>Museum visitors’ opinion dataset</i>
DMP identifier	MuOpinion
Identifier	-
Creators	<i>Joel Lanir</i> , <i>University of Haifa</i>
Contributors	-
Rights Holders	<i>University of Haifa, Museum of Hecht</i>
Description and purpose	Description: Hecht Museum visitor's opinions and comments related to selected exhibits at the museum. Tablets will be placed at selected exhibits at the Hecht museum presenting information on them and asking for visitors to contribute their opinions regarding the exhibits and regarding previous visitor's opinions. Paid participants will be asked to visit the museum, go to the selected exhibits and enter their comments. Regular museum visitors will also be able to enter their comments. Purpose: Classify museum visitors’ opinions according to sentiment and according to topics in dedicated tasks in WP3. Evaluate results of Israelian pilot in WP7.
Data utility	Pilot in WP7
Data types and formats	Qualitative, collected and generated, raw and derivative data. Formats: txt, csv

Table 16 Museum visitors’ opinion dataset

Short name	<i>Haifa case study web application</i>
DMP identifier	HaifaPilot
Identifier	-
Creators	<i>Alan J. Wecker</i> , <i>University of Haifa</i>
Contributors	<i>Joel Lanir</i> , <i>University of Haifa</i>

Rights Holders	<i>University of Haifa (Software reused from other WPs will keep their own rights holders).</i>
Description and purpose	Description: Software to collect data from at most 10 people at a time when visiting the museum, including components developed by other WPs. Purpose: Evaluate results of Israelian pilot in WP7.
Data utility	Pilot in WP7

Table 17 Haifa case study web application

Short name	<i>End users' demographics, preferences and results from activity analysis</i>
DMP identifier	Demog
Identifier	-
Creators	<u>Lily Diaz</u> , Aalto University; <u>Leena Svinhufvud</u> , Design Museum Helsinki; <u>Gautam Vishwanath</u> , Aalto University
Contributors	-
Rights Holders	<i>Aalto University and Design Museum Helsinki</i>
Description and purpose	Description: Responses to semi-structured questionnaires with queries regarding users' preferences and general demographic profile and queries about users' preferences. Part of the data will be collected through questionnaires and part of it through recorded interviews that will be transcribed raw using speech to text software. Purpose: Evaluate results of Finnish pilot in WP7.
Data utility	For other partners of WP7 and case studies.
Data types and formats	Qualitative and quantitative, collected raw and derivative data, including also secondary data to support the creation of derivative data. Formats: docx, txt, csv, xlsx, json, jpg, png, tiff

Table 18 End users' demographics, preferences and results from activity analysis

Short name	<i>Autoethnographic and duoethnographic accounts of material culture</i>
DMP identifier	Ethnog
Identifier	-
Creators	<u>Lily Diaz</u> , Aalto University; <u>Leena Svinhufvud</u> , Design Museum Helsinki; <u>Gautam Vishwanath</u> , Aalto University
Contributors	-
Rights Holders	<i>Aalto University and Design Museum Helsinki</i>
Description and purpose	Description: Responses to questionnaires about culture and narrative texts describing personal relationship to items of culture and historical events. Data are collected through recorded interviews (used in combination with photographs) that will be transcribed raw using speech to text software. Purpose: Evaluate results of Finnish pilot in WP7.
Data utility	For other partners of WP7 and case studies.
Data types and formats	Qualitative and quantitative, collected raw and derivative data, including also secondary data to support the creation of derivative data. Formats: docx, txt, csv, xlsx, json, jpg, png, tiff

Table 19 Autoethnographic and duoethnographic accounts of material culture

Short name	<i>Surveys documenting responses to prototypes</i>
DMP identifier	PortableMu
Identifier	-

Creators	<u>Lily Diaz</u> , <i>Aalto University</i> ; <u>Leena Svinhufvud</u> , <i>Design Museum Helsinki</i> ; <u>Gautam Vishwanath</u> , <i>Aalto University</i>
Contributors	-
Rights Holders	<i>Aalto University and Design Museum Helsinki</i>
Description and purpose	Description: Responses to survey questions in order to understand how a user felt while experiencing the portable museum prototype (see <u>VRpopup</u>), including multiple choices questions on users’ emotions during the use of the prototype. The survey will not collect any personal information or identifier of a participant. The data will be aggregated with other users’ responses and patterns will be identified. Purpose: Evaluate results of Finnish pilot in WP7.
Data utility	For other partners of WP7 and case studies.
Data types and formats	Qualitative and quantitative, collected raw and derivative data, including also secondary data to support the creation of derivative data. Formats: txt, docx, csv, xlsx, json

Table 20 Surveys documenting responses to prototypes

Short name	<u>VR experience of the Pop-up Museum</u>
DMP identifier	VRpopup
Identifier	-
Creators	<u>Lily Diaz</u> , <i>Aalto University</i> ; <u>Leena Svinhufvud</u> , <i>Design Museum Helsinki</i> ; <u>Gautam Vishwanath</u> , <i>Aalto University</i>
Contributors	-
Rights Holders	<i>Aalto University and Design Museum Helsinki</i>
Description and purpose	Description: Users put on the VR headsets and experience the pop-up museum by navigating in the virtual space and interacting with the digital artefacts. Having received consent from the users, their experience/’gameplay’ within the virtual environment will be recorded. User-experience is first designed using role-playing methods such as Bodystorming and Wizard of Oz. This data is captured using a combination of drawing, photography and text. Once the prototype has been produced, users are provided with the headset and the designed experience inside it. Protocol analysis and structured questionnaires are used to elicit descriptive data from users regarding the user experience as well as the usability. Purpose: Evaluate results of Finnish pilot in WP7.
Data utility	For other parts of WP7 and case studies.

Table 21 VR experience of the Pop-up Museum

Short name	<u>User-generated data on the web</u>
DMP identifier	FTMdata
Identifier	-
Creators	Anna Follo, FTM; Giorgia Rochas, FTM; Federica Sesia, FTM
Contributors	-
Rights Holders	<i>Fondazione Torino Musei</i>
Description and purpose	Description: Users’ comments on cultural objects owned by Fondazione Torino Musei (FTM) are collected via a web application developed by WP5 (see <u>IUCitizen</u>). Comments include reactions, drawings, etc. Purpose: Relevant selected comments will be associated to the existing open datasets about FTM collection.
Data utility	FTM and other partners working on WP7 case studies.

Data types and formats Qualitative and quantitative, collected raw and derivative data.
Formats: csv, json, jpeg

Table 22 User-generated data on the web

FAIR Data

Findability

Persistent identifiers and project metadata compliant with H2020 requirements will be provided for every research output by at least one certified repository, namely: Zenodo. Other certified institutional repositories are currently under evaluation to increase the findability of resources. The DOI attributed by Zenodo will be used in publications for referencing underlying data and software and facilitate the harvesting and linking by OpenAIRE.

When depositing in Zenodo, the following **naming conventions** will be used to name new repositories: <SPICE>_<Repository Name>_<Version number>. Version numbers follow specifications detailed in [Semantic Versioning](#). **Keywords** will be provided along with metadata so as to facilitate searchability, and may also include terms from existing classification systems such as the [ACM Classification System](#). Since controlled vocabularies may change significantly within disciplines addressed in the SPICE project, more classification systems will be added in future stages of the project so as to increase findability of resources.

Versioning procedures is ensured by the selected repository. If not differently specified, only the last updated, proofread, and evaluated version of data/software is stored for long-term preservation.

In the following table are listed SPICE research outputs, the additional documentation attached or linked as a source when depositing code or data on Zenodo, the metadata set used to describe the dataset, and the provider of persistent identifier.

DMP identifier	Provision of metadata and documentation	Reuse or development of metadata standards	Persistent identifiers
Workshop	README file	Zenodo Metadata set, CLAAUDIA Metadata set	Zenodo (DOI)
Annotation	Internal documentation	-	-
UM	Javadoc	Zenodo Metadata set	Zenodo (DOI)
Community	README file and user manual	Zenodo Metadata set	Zenodo (DOI)
Recomm	Javadoc	Zenodo Metadata set	Zenodo (DOI)
LDH	README file and interactive documentation	Zenodo Metadata set	Zenodo (DOI)
LDHReg	README file, JSON-LD context	Zenodo Metadata set	Zenodo (DOI)
IGmuseum	README file, JSON-LD context	Zenodo Metadata set	Zenodo (DOI)
IGclassifier	README file, Jupyter notebook	Zenodo Metadata set, AMSActa Metadata set	Zenodo (DOI), AMSActa (DOI)
IUcitizen	README file and interactive documentation	Zenodo Metadata set	Zenodo (DOI)
Reasoner	README file and Javadoc	Zenodo Metadata set	Zenodo (DOI)
Ontologies	Online ontology documentation	Zenodo Metadata set, AMSActa Metadata set	Zenodo (DOI), AMSActa (DOI)

IMMAdata	README file, JSON-LD context	Zenodo Metadata set, OUDA Data Archive ¹ Metadata set	Zenodo (DOI), OUDA (persistent identifier)
IMMAPilot	README file	Zenodo Metadata set	Zenodo (DOI)
MuOpinion	README file	Zenodo Metadata set	Zenodo (DOI)
HaifaPilot	Javadoc	Zenodo Metadata set	Zenodo (DOI)
Demog	Internal docx or README files	Zenodo Metadata set	Zenodo (DOI)
Ethnog	Internal docx or README files	Zenodo Metadata set	Zenodo (DOI)
PortableMu	Internal docx or README files	Zenodo Metadata set	Zenodo (DOI)
VRpopup	Internal docx or README files	Zenodo Metadata set	Zenodo (DOI)
FTMdata	README file	Zenodo Metadata set	Zenodo (DOI)

Table 23 Findability of datasets and software

Accessibility

In the following table are outlined policies for sharing research outputs, including whether data will be fully open (Y) or partially open (P), motivations for closeness and alternative policies, preliminary procedures for data sharing, repositories for dissemination, and required access methods and tools for their reuse.

As specified in the deliverables D9.1, D9.2, D9.3, D9.4, and D9.5 of the Ethics Work Package (WP9), all the data that will be published and made available publicly by any partner in the SPICE consortium will be either **anonymised or pseudonymised**. Specifically, to ensure the privacy of the subject's information, data collected by a member of a consortium will be either anonymised or pseudonymised before being transferred to other stakeholders. More details are provided in Section Ethical aspects.

Preliminary procedures for sharing data also include quality assurance procedures such as data selection, soundproof, copy-editing and proofread of surveyed data.

Institutional repositories and well-known **dissemination platforms** for source code and data will be used along with Zenodo for sharing research outputs and increase their visibility. As outlined in Section [Data Summary](#), data are shared according to standard data formats that allow stakeholders to access and reuse those by means of free-of-charge or open software solutions.

DMP identifier	Open	Motivations and alternatives	Repository	Access methods
Workshop	P	Motivation: Video and recorded materials which cannot be submitted to total anonymization procedures and for which we don't have the user's consent will not be published.	CLAAUDIA (Aalborg University), Zenodo, LDH SPARQL endpoint	Any word processing software
Annotation	N	Motivation: the software will not be open due to commercial opportunities	-	-
UM	Y	N/A	GitHub, Zenodo	Git, Any Java IDE (e.g. ECLIPSE)
Community	Y	N/A	GitHub, Zenodo	Git, Web browser

¹ <https://www.open.ac.uk/library/digital-archive/>

Recomm	Y	N/A	GitHub, Zenodo	Git, Any Java IDE, Web browser
LDH	Y	N/A	GitHub, Zenodo	Git, Web Browser
LDHReg	Y	N/A	Zenodo, LDH SPARQL endpoint	Any Text editor or IDE
IGmuseum	P	Motivation: Users' profile data are not of interest. Users' images cannot be republished without consent. Alternatives: Data will be published in anonymized form. Links to original posts are preserved instead of than images.	GitHub, Zenodo, AMSActa , LDH SPARQL endpoint	Git, Any Text editor or IDE
IGclassifier	Y	N/A	GitHub, Zenodo	Git, Any Text editor or IDE
IUcitizen	Y	N/A	GitLab	Git, Web browser
Reasoner	Y	N/A	GitHub	Git, Web browser
Ontologies	Y	N/A	GitHub, Zenodo, LDH SPARQL endpoint	Git, Web browser, Any Text editor or IDE, Protégé
IMMAdata	P	Motivation: Confidential information released in interviews, high-quality images with copyright restrictions, and data belonging to third parties cannot be disclosed. Alternatives: Data will be anonymized, participant consent, and copyright permission will be asked.	Zenodo, OUDA Data Archive, LDH SPARQL endpoint	Any Text editor or IDE, Web browser
IMMAPilot	Y	N/A	GitHub, Zenodo	Git, Web browser
MuOpinion	P	Motivation: Users' identities are not of interest. Alternatives: Data will be anonymized and/or participants' consent will be asked.	Zenodo, LDH SPARQL endpoint	
HaifaPilot	P	Motivation: Museum contents with copyright restrictions, and individual answers to questions are not shared due to privacy concerns. Alternatives: Partial data will be shared in anonymized form to ensure reproducibility of the evaluation of the application.	GitHub, Zenodo	Git, Web browser
Demog	P	Motivation: sensitive data cannot be shared for privacy concerns. Alternatives: Sensitive information will be omitted in the transcriptions but recorded in the modulated audio recordings.	Zenodo, LDH SPARQL endpoint	Any word processing software
Ethnog	P	See motivations and alternatives of prior dataset <i>Demog</i>	Zenodo, LDH SPARQL endpoint	Any word processing software
PortableMu	P	See motivations and alternatives of prior dataset <i>Demog</i>	Zenodo, LDH SPARQL endpoint	Any word processing software
VRpopup	Y	-	Zenodo	Mozilla Hub
FTMdata	P	Motivation: sensitive data cannot be shared for privacy concerns. Alternatives: Sensitive information will be omitted	Zenodo, LDH SPARQL endpoint	Any word processing software

Table 24 Accessibility of datasets and software

Interoperability

Research outputs will be mostly produced by using **non-proprietary, free-of-charge programming frameworks and stacks of standard technologies**, such as Web languages and Semantic Web

technologies, and by reusing **open-source or free-of-charge software solutions**, so as to ensure their technological interoperability.

To overcome limits due to different data formats and content heterogeneity across datasets, data produced by partners that are relevant to Citizen Curation will *also* be reengineered as Linked Data according to terms defined in the **SPICE Ontologies Network** and made available through dedicated **SPARQL endpoints and APIs** on the **SPICE Linked Data Hub**. The SPICE ontologies are developed by aligning terms to existing stable ontologies so as to ensure their semantic interoperability.

DMP identifier	Standards for interoperability	Ontologies or vocabularies
Workshop		SPICE Ontologies
Annotation	Programming languages: Java, Python Other: SQL, Docker. Web stack (REST, HTTP) Semantic web (JSON-LD)	-
UM	Programming languages: Java	-
Community	Web stack (HTTP, HTML, CSS, JS)	-
Recomm	Programming languages: Java	-
LDH	Web stack (HTTP, HTML, CSS, JS) Semantic Web (RDF, RDFS, OWL2, JSON-LD, SPARQL). Programming languages: PHP, Java, Python. Activity Streams 2.0. Open Digital Rights Language (ODRL 2.2)	-
LDHReg	Semantic Web (RDF, JSON-LD)	DCAT
IGmuseum	Semantic Web (RDF, JSON-LD)	SPICE Ontologies
IGclassifier	Web stack (HTTP, HTML, CSS, JS) Semantic Web (RDF, RDFS, OWL2, JSON-LD, SPARQL) Programming languages: Python.	-
IUcitizen	Web stack (HTTP, HTML, CSS, JS) Semantic Web (RDF, RDFS, OWL2, JSON-LD) Activity Streams 2.0, Open Digital Rights Language (ODRL 2.2). Programming languages: PHP, JavaScript, Python.	-
Reasoner	Semantic Web standards (OWL, RDF, SPARQL) Programming languages: JAVA	-
Ontologies	Semantic Web (RDF, OWL2, SPARQL)	CIDOC-CRM, ArCO
IMMAdata	Semantic Web (RDF, JSON-LD)	SPICE Ontologies
IMMAPilot	Web stack (HTTP, HTML, CSS, JS) Semantic Web (RDF, RDFS, OWL2, JSON-LD, SPARQL) Programming languages: PHP, Java, Python.	-
MuOpinion		SPICE Ontologies
HaifaPilot	Programming languages: Java	-
Demog		SPICE Ontologies
Ethnog		SPICE Ontologies
PortableMu		SPICE Ontologies
VRpopup		
FTMdata		SPICE Ontologies

Table 25 Interoperability of datasets and software

Reusability

Data and software will be licensed under as open as possible **licenses** to foster their reusability. Due to commercial opportunities for some partner, a few software solutions will not be stored for long-term preservation in open repositories.

All data and software are subject to **quality assurance procedures**, some yet to be defined and agreed between partners producing similar datasets.

As aforementioned, **Zenodo** will be the main reference repository for storing all research outputs, along with a few certified institutional repositories, so as to ensure duplication of resources and increase the chances of their long-term availability.

DMP identifier	Licensing and restrictions	Quality assurance processes	Repository for long-term preservation
Workshop	CC-BY 4.0		Zenodo
Annotation	-	Internal peer review. Unit tests and integration tests for continuous development.	-
UM	CC-BY 4.0	Unit and system tests	Zenodo
Community	GPL 3.0	User testing	Zenodo
Recomm	CC-BY 4.0	Unit and system tests	Zenodo
LDH	Apache 2.0	Unit testing, Code review, Issue tracking, User testing	Zenodo
LDHReg	CC-BY-NC 4.0	RDF data validation	Zenodo
IGmuseum	CC-BY 4.0	RDF data validation	Zenodo
IGclassifier	CC-BY 4.0	Issue tracking, User testing	Zenodo, AMSActa
IUCitizen	MIT	Unit testing, Code review, Issue tracking, User testing	Zenodo
Reasoner	Apache 2.0	Unit testing, Code reviewing, Issue tracking	Zenodo
Ontologies	CC-BY 4.0	Ontology consistency validation, Usability	Zenodo, AMSActa
IMMAdat	CC-BY-NC 4.0	RDF data validation	Zenodo
IMMAPilot	Apache 2.0	Unit testing, Pair programming, User tests	Zenodo
MuOpinion	CC-BY 4.0		Zenodo
HaifaPilot	CC-BY 4.0		Zenodo
Demog	CC-BY-NC 4.0		Zenodo
Ethnog	CC-BY-NC 4.0		Zenodo
PortableMu	CC-BY-NC 4.0		Zenodo
VRpopup	CC-BY-NC 4.0		Zenodo
FTMdata	CC-BY 4.0	Data quality is ensured by the application used for collecting data (IUCitizen).	Zenodo

Table 26 Reusability of datasets and software

Allocation of resources

The consortium will use the free-of-charge Zenodo repository for making datasets and software publicly accessible both for dissemination and long-term preservation purposes. This ensures data are safely stored in a certified repository and easily harvestable by EU services such as OpenAIRE.

Every member of the consortium is responsible for creating their own account on Zenodo and to deposit data and software according to procedures stated in Section “FAIR Data”.

As for the publications, all the findings derived from research data must be published in scientific journals that allow green open access or golden open access. In the case of gold open access, costs related to open access will be claimed as part of the Horizon 2020 grant, compliantly with the budget of each partner of the SPICE consortium.

Data security and reuse

All documents relevant to the project are stored in a dedicated **GDPR-compliant remote sharing platform (Microsoft 365)** provided by the University of Bologna, which bought a license that is available to any member of the institution and to external guests that have a Microsoft account. In this private storage area, access is granted only to project partners after requesting permission. In case sensitive data are here stored, these are uploaded only after anonymization procedures, so as to ensure no misuse by other partners is possible. Data recovery and secure storage is guaranteed by the service provider.

Moreover, the **SPICE Linked Data Hub (LDH)** developed in WP4 will allow partners to upload their datasets and have full control over the definition of rules for data reuse, including the definition of copyright restrictions and licenses to any asset they manage. Specifically, the SPICE LDH implements an access control mechanism that allows dataset owners to define their own policies for accessing the dataset. Users can subscribe to the dataset and the owner can grant read/write access to it. This mechanism revolves around the concept of key. A key is a sequence of characters (i.e., a password) that enables a user to perform an operation (i.e., read or write) on a given dataset on the basis of the policy configured by the dataset owner. Keys can be:

- *Read only.* The user can only read the dataset;
- *Write-only.* Dataset owners may wish to grant write access to users or applications that need to contribute data to the dataset without having read access to other data potentially submitted by other users;
- *Read and Write.*

Dataset owners can revoke access to specific keys at any time. Where access policy has changed or an application may have become compromised, there are scenarios where a key owner may wish to remove their own access to a dataset or a dataset owner may want to disable specific user or key access.

Moreover, the LDH will provide a provenance layer (which is currently under development) which is meant to streamline the collection and delivery of provenance metadata related to cultural heritage objects (CHO). The requirements of the provenance layer are outlined in the deliverable D4.5.

Finally, LDH also ensures data recovery and data backup on a regular basis. The data will be stored in the LDH for the duration of the project. The dataset owner can then decide whether to keep or delete the data. Long-term secure storage of final versions of datasets and software is guaranteed for all aforementioned research output by **Zenodo**.

Ethical aspects

All the aspects related to ethical issues are addressed in the deliverables related to Work Package 9 (WP9). In particular, currently, the following deliverables have been already submitted to the European Commission for review:

- D9.1 – POPD – H – Requirement No.3
- D9.2 – POPD – Requirement No.11
- D9.3 – POPD – Requirement No.12
- D9.4 – POPD – Requirement No.13
- D9.5 – NEC – Requirement No. 16
- D9.6 – POPD – Requirement No. 17

We provide a summary of the ethical aspects discussed in those deliverables as follows.

- *Personal Data.* The definition of personal data is set out in art. 4 of the GDPR, and it states as follows: ‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an

identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

- The SPICE consortium will limit the collection of personal data that allows the identification of the data subject/research participant either directly or indirectly.
- The Data Controller (i.e. the member of the consortium who is charge of collecting data) should assure that personal data collected cannot be linked to or singled out a data subject, either by avoiding the collection of specific personal data or by assuring the deletion of identifiers (such as name, identification number, location data, etc.). All the personal data that will be collected by a member of a consortium will be necessarily either anonymised or pseudonymised before being transferred to entities (including other members of the project consortium) different from the one in charge of processing such personal data.
- As stated in SPICE Deliverable D9.2, where the collection of personal data is desirable, it will be limited to the scope of the partner acting as Data Controller and with the sole purpose of allowing the data owners to act on their rights with relation to GDPR.
- It is a recommended policy to use one or more pseudonymisation techniques to enable the exchange of data between partners and where it is not necessary to know the identity of the data subject.
- Anonymisation and pseudonymisation techniques adopted include:
 - *Data Suppression.* Removing identifier references such as data subjects' name, address, location data, online identifier, postcode information.
 - *Data Generalisation.* Modify the precision of the gathered variables, where possible and according to the type of data, represent data using scales. For example, age range (0-3, 4-6), partial information such as year of birth rather than full date.
 - *Data Encryption.* A technique that uses a secret key to obtain pseudonyms from an identifier, it uses an encryption algorithm.
 - *Data Tokenisation.* A non-mathematical approach that generates randomly values that replace data identifiers with non-sensitive information.
 - *Data swapping.* Rearranging records so that they do not match the initial information.
 - *Data perturbation.* changing the original data by using rounding methods and random noise.