



D1.2

Data Management Plan

2016-08-29

Yannick Naudet, Ioanna Lykourantzou, Aggeliki Antoniou, Daphne Kyriaki-Manessi,
Kalliopi Kontiza, Joseph Padfield, Catherine Jones, Andreas Vlachidis, Martín López-Nores,
Alberto Gil-Solla

The CROSSCULT project participates in the Open Research Data Pilot, which aims to make the research data generated by selected Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access. Open data is data that is free to access, reuse, repurpose, and redistribute. This document contains the initial version of the Data Management Plan (DMP) of CROSSCULT, describing how the research data collected or generated will be handled during and after the project. This will be revised at mid-term and at the end of the project.



www.CROSSCULT.eu

Project acronym	CROSSCULT
Full title	CROSSCULT: Empowering reuse of digital cultural heritage in context-aware crosscuts of European history
Grant agreement number	693150
Funding scheme	Innovation Action (IA)
Work programme topic	H2020-REFLECTIVE-2014-2015/REFLECTIVE-6-2015: Innovation ecosystems of digital cultural assets
Project start date	2016-03-01
Project duration	36 months

Workpackage 1	Management
Deliverable lead organisation	University of Vigo
Authors	Yannick Naudet, Ioanna Lykourantzou (LIST), Aggeliki Antoniou (UoP), Daphne Kyriaki-Manessi (TEI-A), Kalliopi Kontiza, Joseph Padfield (NG), Catherine Jones (UL), Andreas Vlachidis (UCL), Martín López-Nores, Alberto Gil-Solla (UVigo)
Reviewers	Evgenia Vasilakaki (TEI-A) George Lepouras (UoP)
Version	1.0
Status	Final
Dissemination level	PU: Public
Due date	M6 (2016-08-31)
Delivery date	2016-08-29

Table of Contents

1. Introduction.....	6
2. Selection of data repositories.....	8
3. General datasets.....	9
4. Pilot 1 datasets	14
5. Pilot 2 datasets	27
6. Pilot 3 datasets	33
7. Pilot 4 datasets	38
8. DMP Updates.....	42

List of Tables

Table 1. Dataset information template	6
Table 2. Upper-level ontology	9
Table 3. Literals of relevant dates	10
Table 4. Venue infrastructure and collections	11
Table 5. Project publications and bibliography	13
Table 6. Pilot 1 NG collection ontology extensions and literals	14
Table 7. Pilot 1 user profiles.....	17
Table 8. Pilot 1 user-generated content.....	18
Table 9. Pilot 1 NG routes and tours	21
Table 10. Pilot 1 multimedia images	22
Table 11. Pilot 1 user tracking and observations	24
Table 12. Pilot 1 game scenarios	25
Table 13. Pilot 2 ontology extensions and literals.....	27
Table 14. Pilot 2 visitor profiles	28
Table 15. Pilot 2 multimedia contents	29
Table 16. Pilot 2 game scenarios	30
Table 17. Pilot 2 visitor interactions and staff observations	31
Table 18. Pilot 2 interview transcripts.....	32
Table 19. Pilot 3 ontology extensions and literals.....	33
Table 20. Pilot 3 visitor profiles	34
Table 21. Pilot 3 multimedia contents	35
Table 22. Pilot 3 visitor interactions and staff observations	36
Table 23. Pilot 3 interview transcripts.....	37
Table 24. Pilot 4 ontology extensions and literals.....	38
Table 25. Pilot 4 geolocated object discovery sequences	38

Table 26. Pilot 4 reflections on History	39
Table 27. Pilot 4 collective memory stories	39
Table 28. Pilot 4 questionnaire responses	40
Table 29. Pilot 4 interview transcripts.....	41

1. Introduction

CROSSCULT has been chosen to participate in the Open Research Data Pilot (<https://www.openaire.eu/opendatapilot>), which aims to make the research data generated by selected Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access.

Open data is data that is free to access, reuse, repurpose, and redistribute. The basic requirement from OpenAIRE is to deposit research data in a repository where they will be findable and accessible for others. The partners are not expected to share sensitive data or breach any IPR agreements with industrial partners. It is not necessary to deposit all the data generated during the project either; only data that underpins published research findings and/or has longer-term value. This applies to (i) the data and metadata needed to validate results in scientific publications and (ii) other curated and/or raw data and metadata that may be required for validation purposes or with reuse value. Examples include anonymised user profiles, contextual information, statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.

This deliverable is the initial version of the Data Management Plan (DMP). According to the Guidelines on Data Management in Horizon 2020 (v2.0):

- The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.
- The DMP is not a fixed document, but evolves during the lifespan of the project. The initial version will be updated at least during mid-term and towards the end of the project taking into account the final reviews in order to fine-tune it to the data generated and the uses identified by the consortium.
- The DMP should describe all the datasets generated or collected during the project according to the template provided in Table 1, reflecting the current status of reflection within the consortium about the data that will be produced.

Table 1. Dataset information template

<Dataset reference>	<Dataset name>
<i>Dataset description</i>	Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
<i>Standards and metadata</i>	Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created, in order to enable data to be found and understood. Metadata should reply to such questions as: <ul style="list-style-type: none"> • What is the data about?

<Dataset reference>	<Dataset name>
	<ul style="list-style-type: none"> • Who created it and why? • In what forms it is available? • What is the inner structure of folders and files organisation? • What is the naming scheme to manage different versions of the dataset?
<i>Data sharing</i>	<p>Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling reuse, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).</p> <p>In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.</p>

This document contains the tables corresponding to the datasets identified from the work carried out in WP2, leading to the refinement of the CROSSCULT pilot specifications. The naming scheme will be based on the guidelines specified in the “*Naming conventions and document versioning*” section (4.2.2.3) of Deliverable D1.1 (“*Quality and risk management plan*”). There are two types of datasets:

- General datasets of the project (e.g. the upper-level ontology) will get names starting with “*CROSSCULT-DS-*”.
- Pilot-specific datasets will get names starting with “*CROSSCULT-DS-pilot_x_*”, “*x*” being the pilot number used for a reference in the CROSSCULT grant agreement.

2. Selection of data repositories

The selection of a data repository to serve the project's needs was addressed during the General Assembly of May 2016 in Luxembourg, where a number of options were identified:

- On the one hand, it was noted that TEI-A and UNIPD have established data management facilities of their own.
- On the other hand, there were a number of external repositories listed at the *Registry of Research Data Repositories* (<http://www.re3data.org/>), as well as recommendations from the *Open Access Infrastructure for Research in Europe* (OpenAIRE) to bear in mind.

During the first half of June 2016, the discussion of the Steering Committee came down to the following options:

- 1) Using zenodo (<http://zenodo.org/>), an open repository created by OpenAIRE and CERN to provide a place for researchers to deposit any kind of material (data, publications, etc.). It allows uploading files up to 2GB and offers flexible licensing schemes. This is a popular choice in other H2020 projects.
- 2) Using HAL (<https://hal.archives-ouvertes.fr/>) is an open archive where authors can deposit scholarly documents from all academic fields, enjoying a good position in the international web repository rankings. HAL is run by a French computing centre, which is part of the *Centre National de la Recherche Scientifique* (CNRS), and supported by other French institutions like INRIA.
- 3) Using the institutional repository of TEI-A or UNIPD.
- 4) Using a combination of external and institutional repositories with automated replication.
- 5) If the preceding option were not feasible, using a combination of external and institutional repositories with manual replication.

After checking the possibilities of the institutional repositories for automatic replication, the final decision was made for option 4: zenodo with a mirror managed by TEI-A located at <http://hypatia.teiath.gr/>. The latter is also linked to a local harvester called "open archives" and through this to Google Scholar and very soon to Europeana, which ensures extra dissemination at very little cost. This decision applies to all the datasets described in subsequent pages of this document, unless otherwise specified in the corresponding tables.

Also in June 2016, it was decided to use Mendeley (<http://www.mendeley.com>) as a tool for reference management for the whole CROSSCULT consortium.

3. General datasets

The general datasets identified hitherto refer to two ontologies that may be used to support reasoning, adaptation and personalisation procedures in any of the pilots (namely an upper-level ontology for cultural heritage and an ontology of relevant dates that may underpin context awareness features), plus a repository of information about venues involved in CROSSCULT experiences (including those to come from Living Lab activities) and the repository of project publications and relevant bibliographical references.

Table 2. Upper-level ontology

CROSSCULT-DS-Upper_Level_Ontology	Upper-level ontology
<i>Dataset description</i>	<p>This will be a single generic upper-level ontology for cultural heritage, supporting many computational processes of the CROSSCULT platform. It will be a common umbrella for the vocabularies, ontologies and other metadata models that are currently used by the different venues where the pilots will take place.</p> <p>The ontology will rely on standard Semantic Web languages and standard ontology models for cultural heritage such as CIDOC-CRM, an international standard for the exchange of cultural heritage information, commonly used by several cultural heritage organisations like the National Gallery, the British Museum and others. When necessary the ontology will consider official CRM extensions: e.g. CRMsci (<i>Scientific Observation Model</i>), CRMdig (<i>Digital Provenance Model</i>), CRMgeo (<i>Spatiotemporal Refinement</i>). Further augmentations of the ontology with standard semantic web ontologies, such as SKOS and FOAF could be introduced to complement the semantics of CRM.</p> <p>This upper-level ontological layer will facilitate the interoperability and information exchange needs between the different venues, but also the exploitation of external data sources such as Europeana, the Getty Vocabularies, the British Museum digital resources and other linked data vocabularies following the same metadata standards.</p> <p>This dataset will be open, and could be used (and, enhanced) by researchers and experts worldwide for any purposes. It will be delivered in first version as Deliverable 2.3 of the project (<i>“Upper-level cultural heritage data structure & ontology - First version”</i>) by October 2016, and then in final version as Deliverable 2.5 (<i>“Upper-level cultural heritage data structure & ontology”</i>) by March 2018.</p>
<i>Standards and metadata</i>	<p>The ontology will be created using the OWL standard. The size will be manageable in terms of storage requirements. The estimation is that the ontology could get as big as a few thousand nodes and tens of thousands of relations to external concepts through semantic properties.</p>
<i>Data sharing</i>	<p>The ontology will be shared with any interested researcher through the public general project repositories. Any software tool supporting the OWL standard may be used to visualise the data.</p> <p>All the data belonging to the dataset will correspond to public</p>

	<p>information. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access both data and metadata online with no restrictions at all.</p> <p>The project will take into account the Creative Commons policies as described by the organisation in regards to access, reuse, reproduction and distribution.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The ontology will be stored and made available to worldwide researchers and experts through the project repositories (see Section 2). Since these repositories are free, no costs are foreseen for long-term storage.</p>

Table 3. Literals of relevant dates

CROSSCULT-DS- Relevant_Dates_ Literals	Relevant dates literals
<i>Dataset description</i>	<p>This will be a collection of literals of significant dates around Europe (“<i>international days of</i>”, “<i>commemorative dates</i>”, “<i>national days of</i>”, “<i>birth dates of</i>”) that may provide propitious contexts to explain interconnections among different venues. It will specify relevant dates, including the description of the underlying significance, and properties that may be used to relate it to other dates, to historical events or characters, to the places where they are relevant, etc. These data can be used to feed reasoning processes that will discover links between venues that are worth highlighting in specific dates, and thereby make proposals about interesting experiences involving several venues. The origin of the data will be decided in a later stage, considering existing alternatives that could serve “<i>as-is</i>” for the task or that could be conveniently modified. Relevant input has been found in the following links:</p> <ul style="list-style-type: none"> • United Nations’ International Days: http://www.un.org/en/sections/observances/international-days/ • UNESCO International Days: http://en.unesco.org/celebrations/international-days • List of international observances: https://en.wikipedia.org/wiki/List_of_International_observances • List of National Days: https://en.wikipedia.org/wiki/National_Day • List of Flag Days: https://en.wikipedia.org/wiki/Flag_Day • List of Liberation Days: https://en.wikipedia.org/wiki/Liberation_Day • List of commemorative days: https://en.wikipedia.org/wiki/List_of_commemorative_days • List of awareness days, weeks and months: https://en.wikipedia.org/wiki/List_of_awareness_days • Dates related to relevant characters following the “<i>birthDate</i>” property of DBpedia: http://dbpedia.org/ontology/birthDate <p>This dataset could be used in research publications, but only as necessary input to illustrate whichever reasoning mechanisms. It may be used (and, possibly, enhanced) by researchers worldwide in any experiments that would consider dates as a relevant dimension of</p>

	context in personalisation or recommendation systems.
<i>Standards and metadata</i>	This dataset will be created using the OWL standard, with date formats obeying ISO 8601 specifications. The size of the ontology will be small in terms of storage requirements, since it is not expected that there will be more than a few observations per day of the year, and the data is text only. The estimation is that the ontology could get as big as a few thousand nodes and tens of thousands of relations to external concepts through semantic properties.
<i>Data sharing</i>	<p>This dataset will be shared with any interested researcher through the public general project repository. Any software tool supporting the OWL standard may be used to visualise the data.</p> <p>All the data belonging to the dataset will correspond to public information about important dates, their significance, geographical sites associated and relationships among them. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access both data and metadata online with no restrictions.</p> <p>The project will take into account the Creative Commons policies as described by the organisation in regards to access, reuse, reproduction and distribution.</p>
<i>Archiving and preservation (including storage and backup)</i>	The dataset will be available to worldwide researchers through the project repositories (see Section 2). No costs are foreseen for long-term storage.

Table 4. Venue infrastructure and collections

CROSSCULT-DS-Venue_Infrastructure_Collections	Venue infrastructure and collections
<i>Dataset description</i>	<p>This dataset will include the description of the venues involved in CROSSCULT experiences, their infrastructure and exhibition characteristics. Initially, it will contain data of the National Gallery, the four venues identified in the proposal for pilot 2 (the Roman healing spa of Lugo, the Roman healing spa and the city of Chaves, the thermo-mineral site of Montegrotto Terme and the archaeological site of Epidaurus) and for pilot 3 (the Archaeological Museum of Tripolis). For the rooms that contain paintings that will be used in the scenario of pilot 1, a sub-dataset will also be created on a room-by-room (or even wall-by-wall) basis. This dataset will also relate to the Group details captured in the <i>CROSSCULT-DS-pilot_1_User_Generated_Content</i> dataset. Data about new venues, involved through the CROSSCULT Living Lab will be added in due time.</p> <p>The information will cover the description of the facilities, including area, maps, number of rooms, maximum capacity, opening hours and dates, availability of toilets, cafeteria, bookshop, souvenir shop, ... plus metadata about the collections in exhibit and appropriate keywords or semantic tags to drive the identification of relevant dates, topics to link</p>

	<p>several venues, or whichever expected outcomes of the computational mechanisms of the CROSSCULT platform. By its own nature, it will be important data to justify the experiment approaches and stages in any publication derived from the project.</p> <p>Collection/Room details:</p> <ul style="list-style-type: none"> • Collection/room ID: Unique reference, not editable by the user. • Collection/room parent ID: Used to identify how a given room relates, hierarchically to a collection of rooms. • Collection/room title: Textual name for the collection and each room. • Collection/room description: Short textual description of the collection and optional for each room. The room descriptions could be displayed as virtual room information panels. • Collection/room dimensions: X, Y, Z, orientation, positions, etc. The minimal amount of dimensional information required to define the size and relative location of each defined room. <p>This dataset may be used by worldwide researchers to organise additional experiments in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc.</p>
<i>Standards and metadata</i>	<p>The dataset items related to the objective data will be collected from descriptions of the venues facilities provided by the staff. The dataset will be structured as a collection of files, one per venue. The project will use CIDOC CRM, expanding the upper-level ontology as required. The description of a location will be included in the upper-level ontology.</p>
<i>Data sharing</i>	<p>There are IPR and security issues for some of the data in this repository (e.g. maps), which still need to be addressed. If these can be cleared, this dataset can be made available to any interested researcher. If not, then parts of it will need to remain private.</p> <p>The format of the files (XML) will allow reuse and further data enrichment. The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing.</p> <p>Generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible. Once cleared, all the information files belonging to this dataset will be deposited in an appropriate data repository to facilitate reuse.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The information volume of this dataset will vary depending on the size of each venue. No costs are foreseen for long-term storage.</p> <p>At the end of the project, final versions of the dataset will be archived as part of the ongoing NG backup procedures. Where appropriate, the data held within this dataset will also be incorporated into internal NG systems.</p>

Table 5. Project publications and bibliography

CROSSCULT-DS-Publications_Bibliography	Project publications and bibliography
<i>Dataset description</i>	This will be a collection of project publications and relevant bibliographical references in the areas of interest for CROSSCULT. Publications will be stored with links to the relevant repositories of research data.
<i>Standards and metadata</i>	Publications will be kept in PDF format; bibliographical references will be stored in .bib format.
<i>Data sharing</i>	The files will be shared with any interested researcher through the public general project repositories. Publications will also be listed on crosscult.eu, the designated site for the project with links to the institutional repositories where the publications are stored.
<i>Archiving and preservation (including storage and backup)</i>	The files will be available to worldwide researchers through the project repositories (see Section 2). No costs are foreseen for long-term storage.

4. Pilot 1 datasets

Table 6. Pilot 1 NG collection ontology extensions and literals

<p>CROSSCULT-DS-pilot_1_NG_Collection_Ontology_Extensions</p>	<p>Pilot 1 National Gallery collection ontology extensions and literals</p>
<p>CROSSCULT-DS-pilot_1_NG_Collection_Literals</p>	
<p><i>Dataset description</i></p>	<p>These datasets will be used to model the ontology of the paintings of the National Gallery collection. These datasets will be composed of the extensions of the upper-level ontology used to describe the types and relationships required to capture the NG collection and the actual literal data related to the collection.</p> <p>The core (Tombstone) data will be retrieved from NG collection information held in the National Gallery Collection Management System (CMS) TMS (The Museum System™) and will be presented to the CROSSCULT app via a new API.</p> <p>Painting details:</p> <ul style="list-style-type: none"> • Painting inventory ID: Accession Number, unique painting ID. • Painting date(s): relevant dates for the painting, including date of production, dates of exhibitions and modifications, etc. • Artist(s): The name of the artist or artist involved in the production of the painting. This will also include details relating to unknown groups of related artists, such as “<i>Workshop of ...</i>”, “<i>Follower of ...</i>”, etc. • Group: Indicates if a given painting is part of a defined group of paintings. The paintings in these groups are normally directly physically related rather than of a similar type. For example paintings that used to be part of the same altarpiece, paintings that were all created as part of one installation, double sided paintings, etc. • Painting title: Full title of a painting. Additional alternative titles may also be available; a shortened version of the title will also be available. • Medium and support: Short terms used to describe the main key materials used to create a given painting, for example “<i>Oil on Canvas</i>”. • Painting dimensions: The physical height and widths of a given painting in centimetres. • Credit line: Were available, details of specific acquisitions credits,

	<p>including the name and date of a given bequest. This can include details of more than one event and date.</p> <ul style="list-style-type: none"> • Public locations: The name or number of the specific Gallery in which the painting is held. All paintings that are not on display are given the generic location of “<i>Not on display</i>”. This will be related to a specific Room ID from the <i>CROSSCULT-DS-pilot_1_Venue_Profile_Rooms</i> dataset. • Relative wall position: It is intended that CROSSCULT will allow the basic room data to be augmented to include specific wall and position based information, subject to the development of appropriate tools and availability of data. • Description: Short textural description of the painting and its history, if available. • Inscription summary: Textural details describing the presence and locations of any specific marks, signatures, dates or more general inscriptions noted on a given painting. • Classifications and keywords: General type and grouping classification terms, along with more general subject matter related keywords. <p>Artist details:</p> <ul style="list-style-type: none"> • Unique Artist ID. • Artist name: Where possible including known variations and translations of these names. • Artist date(s): Generally the date of birth and death of an artist, but possibly dates relating to when they were known to be alive, active or when their work was documented. • Short artist’s biography, where available. <p>This is the core of the paintings dataset that will be used to define the collection and its inter-relationships. This dataset will be further enriched with appropriate internal NG resources of contextual information including:</p> <ul style="list-style-type: none"> • Technical information: narrative text descriptions for supports, media, pigments and techniques. • Other people: related artists, patrons, former owners, sitters – names, dates, biographies. • Places: The names and coordinates of locations related to and or depicted in paintings. This will not be available for all paintings. • Additional events: Related artistic periods and events, including those depicted in paintings- names, dates, descriptions. • Other thematic groupings: descriptions only for genres, stories, styles. <p>These datasets will then be augmented by related concepts such as geographical data and related vocabulary terms and connected to</p>
--	--

	<p>external sources of information (Wikidata, Wikipedia) and other museums APIs (BM, Europeana). The extensive data in Raphael Research Resource (http://cima.ng-london.org.uk/documentation) available via a SPARQL end-point (http://rdf.ng-london.org.uk/raphael) will also be employed as appropriate.</p> <p>The ontology may be used for further semantic related projects to help define and describe similar collection and as part of more general semantic research.</p> <p>The literal dataset may be used in further research to examine the relationships captured in the dataset, further explore the collections connections to the broader world history, further explore the visualisation of complex cultural heritage datasets and or as an example for how other collections might be described.</p>
<p><i>Standards and metadata</i></p>	<p>These datasets will be composed of the ontology(ies) used to describe the types and relationships required to capture the National Gallery Collection and the actual literal data related to the collection. It will be created by NG staff along with assistants from other CROSSCULT researchers, as part of other related NG projects and an ongoing program within the NG to develop and enhance its collection information.</p> <p>The final literal dataset will take the form of structured XML files, with the information modelled in OWL. The types and relationships captured within these files will be documented within the overall ontology, which will also take the form of structured XML. The overall ontology will be based on the CIDOC-CRM cultural heritage standard, along with other related ontologies and enhancements that have been selected and or developed within the CROSSCULT project.</p> <p>Specifically, other thesauri and controlled vocabularies will be used to connect available pieces of data together and provide further information on the collected data (artists, names, materials, geographic locations and object details). For example:</p> <ul style="list-style-type: none"> • The Getty Vocabularies: <ul style="list-style-type: none"> ○ Art & Architecture Thesaurus (AAT) ® ○ Cultural Objects Name Authority (CONA) ® ○ Getty Thesaurus of Geographic Names (TGN) ® ○ Union List of Artist Names (ULAN) ® • The Conservation & Art Materials Encyclopaedia Online (Cameo)
<p><i>Data sharing</i></p>	<p>Much of the data belonging to the literals dataset and the related fields in TMS are already public on the NG website and are in common use. The core data will be made available in accordance to the agreement signed by the Consortium.</p> <p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p>

	<p>Further generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be deposited in an appropriate data repository to facilitate reuse.</p> <p>Access to any augmentation of the core dataset created outside the CROSSCULT app will be as open as possible but specific agreements may well be required (see Consortium Agreement).</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The ontology extensions will be added to the recommended project repository. The literals will encompass dynamically updated information, which will be presented through a documented API, hosted at the National Gallery. A description of the API including the documentation will be added to the recommended project repository.</p> <p>The NG already conducts backups every night to tape on site in a server room facility and the backup(s) are also stored offsite with an external organisation. Some of the costs for this will be covered by CROSSCULT during the project. At the end of the project final versions of the dataset will be archived as part of the ongoing NG backup procedure.</p> <p>Created data that augment the core of the NG paintings dataset will be folded into the NG Information System. When appropriate, this data would also be integrated into the Collection and Information Repositories and continue to be used within future projects of the NG.</p>

Table 7. Pilot 1 user profiles

CROSSCULT-DS-pilot_1_User_Profiles	Pilot 1 user profiles
<i>Dataset description</i>	<p>This dataset will include all the information entered by the users, visitors participating in the experiments of pilot 1, and the data calculated or determined by the CROSSCULT platform, to describe and categorise users and their interests, activities, preferences and interaction with the collection and CROSSCULT app.</p> <p>It comprises anonymised information about the users who will participate in the Gallery visits, including psycho-demographic data. Additionally, for users who give their permission, information can be automatically extracted/retrieved from social networks to infer further interesting profile traits. The exact make-up of the fields included in this dataset will be determined as part of the work carried out within CROSSCULT.</p> <p>This dataset could be used by worldwide researchers to run experiments with similar or different profiles in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc.</p> <p>This dataset may underpin research publications derived from the project, as the characteristics of the users are determinant when evaluating the conclusions of the project experiments.</p>

<p><i>Standards and metadata</i></p>	<p>The dataset will take the form of structured XML files initially and it will then be modelled using primarily the FOAF ontology. FOAF will be completed with other existing ontologies or datasets to build precise geo-located and context-aware user profiles to be used by recommender systems, like, e.g., the Weighted Interests ontology, the Geonames dataset, OWL Time, or other ontologies for representing user activities, abilities, preferences or context. A complete list of reused ontologies will be given once the user ontology is built in the CROSSCULT project.</p> <p>FOAF is an ontology mostly focused on people's existence in the virtual world, with many properties related to online activity or identity. FOAF is particularly well suited for describing people on Web-based Social platforms.</p> <p>The ontology is essentially compatible with versions of OWL.</p> <p>Since FOAF is a rather small ontology, other user/people related ontologies could possibly be used such as the UPOS (<i>User Profile Ontology with Situation-Dependent Preferences Support</i>) and the CIDOC-CRM person entity.</p>
<p><i>Data sharing</i></p>	<p>All the data belonging to this dataset will be anonymised data, unless the information has already been specifically made public by the user(s). They will have no privacy concerns and can be made available to any interested researcher.</p> <p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p> <p>The anonymised user profile metadata needed to validate results in scientific publications and underpin published research findings, as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible. Tests will be conducted to see if anonymised data can still be used to trace back to persons.</p>
<p><i>Archiving and preservation (including storage and backup)</i></p>	<p>All the information files belonging to this dataset will be deposited in the project repositories to facilitate reuse. The archived information will consist of XML files containing the findings obtained through information retrieval.</p> <p>The NG already conducts backups every night to tape on site in a server room facility and the backup(s) are also stored offsite with an external organisation. Some of the costs for this will be covered by CROSSCULT during the project. At the end of the project final versions of the dataset will be archived as part of the ongoing NG backup procedure.</p>

Table 8. Pilot 1 user-generated content

CROSSCULT-DS-pilot_1_User_Generated_Content	Pilot 1 user-generated content
<i>Dataset description</i>	<p>This dataset includes the contributions of content created from the users participating in the experiments using the CROSSCULT app. Its structure will be very similar to the <i>CROSSCULT-DS-pilot_1_NG_Collection_Literals</i> dataset as both will share the same ontology.</p> <p>This content will represent the users' reinterpretation of the NG collection information, allowing them to reflect on their experiences. This will include data literals from any subgroups of the NG collection created by user input, preferences and searches.</p> <p>While interacting with the CROSSCULT app and the content of the NG collection, users will be able to define multiple groups and add unique descriptions of the groups and the objects in them, thus allowing different versions of generated content that will adding to their reinterpretation the collection.</p> <p>User-defined painting groups:</p> <ul style="list-style-type: none"> • Group ID: Unique reference – determined by the CROSSCULT app not editable by the user. • Related room ID: If a group is being modelled on existing NG room(s) then all of the dimensional details can be referenced. • Group parent ID: (Optional) Allow the users to create hierarchical groups. This functionality will allow users connect groups together in the form of virtual rooms to create their own exhibitions. • Group title: Textural name for each given groups. Each user will be given a default “<i>Favourites</i>” group to add paintings to, but will have the option to re-name this and or create additional groups. • Group description: Short textual description of a given group. • Group dimensions: (Optional) This information will only be required if the user actually wants to arrange the paintings in some kind of virtual space. The default option for any groups will be a simple collection of paintings the users wants to group together. If used, the minimal amount of dimensional information required by the CROSSCULT app to define the size and relative location of each defined virtual space. This field will need to be broken down into all of the individual values required. <p>Painting details:</p> <ul style="list-style-type: none"> • Relative painting ID: Each painting in each group can be described different by each user, so each instance of its use in any group will need to be uniquely identified. • Painting ID: Reference to the original Painting ID. • Group ID: Reference. • Painting title: Optional alternative title, default value being a reference to original.

	<ul style="list-style-type: none"> • Description: Optional alternative description, default value being a reference to original. • Classifications and keywords: Optional terms, including or excluding the originals. • Relative locations: (Optional) Default values taken from the actual location of the painting, but editable if required in relation to virtual spaces. • Options to update all of the other painting fields, except the IDs, could also be provided but are less likely to be used. <p>Artist Details:</p> <ul style="list-style-type: none"> • Artist ID: Reference to the original artist. • Alternative artist name. • Alternative artist date(s). • Alternative short artist's biography. <p>Alternative or additional versions of all of the additional fields described in the <i>CROSSCULT-DS-pilot_1_NG_Collection_Literals</i> dataset may also be contained within this dataset.</p> <ul style="list-style-type: none"> • Technical information: narrative texts descriptions for supports, media, pigments and techniques. • Other people: related artists, patrons, former owners, sitters – names, dates, biographies. • Places: The names and coordinates of locations related to and or depicted in paintings. This will not be available for all paintings. • Additional Events: Related artistic periods and events, including those depicted in paintings- names, dates, descriptions. • Other thematic groupings: descriptions only for genres, stories, styles, etc. <p>The literal dataset may be used in further research to examine the relationships captured in the dataset, further explore the collections connections to the broader world history, further explore the visualisation of complex cultural heritage datasets and or as an example for how other collections might be described.</p> <p>This dataset may underpin research publications derived from the project, as the characteristics of the users are determinant when evaluating the conclusions of the experience.</p>
<p><i>Standards and metadata</i></p>	<p>This dataset will be composed of the actual literal data used to describe the collection and their experiences by the users. It will be created by the CROSSCULT app users, along with possible augmentation and inference carried out by CROSSCULT researchers.</p> <p>The final literal dataset will take the form of structured XML files, with the information modelled in OWL. The types and relationships captured within these files will be documented within the <i>CROSSCULT-DS-</i></p>

	<p><i>pilot_1_NG_Collection_Literals</i> ontology.</p> <p>As with the <i>CROSSCULT-DS-pilot_1_NG_Collection_Literals</i> dataset other thesauri and controlled vocabularies will be used to connect available pieces of data together and provide further information on the collected data.</p>
<i>Data sharing</i>	<p>All the data belonging to this dataset will be anonymised data, unless the information has already been specifically made public by the user(s). They will have no privacy concerns and can be made available to any interested researcher.</p> <p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p> <p>User generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be deposited in the project repositories to facilitate reuse. The archived information will consist of XML documentation files.</p> <p>The NG already conducts backups every night to tape on site in a server room facility and the backup(s) are also stored offsite with an external organisation. Some of the costs for this will be covered by CROSSCULT during the project. At the end of the project final versions of the dataset will be archived as part of the ongoing NG backup procedure.</p>

Table 9. Pilot 1 NG routes and tours

CROSSCULT-DS-pilot_1_NG_Routes_Tours	Pilot 1 NG routes and tours
<i>Dataset description</i>	<p>This dataset will include the data collected from paintings location in combination with records of tracking location of users participating in the experiments using the CROSSCULT app. It will begin with a small set of predefined, suggested, NG tours, based on the available data and the defined CROSSCULT pilot 1 reflection points. The dataset will then be augmented by user's recommended tours created during the profiling and recommendation processes, records of user visits, and the creation of new user defined tours.</p> <p>This dataset will include people's visits and activities in a reusable form that can be used by other users, or the system. Some degree of review or scoring could be included, based of automated monitoring of their use and user defined values.</p> <p>This dataset could be used by worldwide researchers to run experiments with similar or different profiles systems and as a method of testing the recommender systems, comparing recommendations with actual</p>

	<p>activities. The data could then be used to improve the recommendations systems within any future project.</p> <p>This will be a dynamic set as users may add new tags and keywords at any time, maybe due to new inputs or changes to the physical layout of the collection. By its own nature, it will be important data to justify the experiment approaches and stages in any publication derived from the project. Tests will be conducted to see if anonymised data can still be used to trace back to persons</p>
<i>Standards and metadata</i>	<p>This dataset will be a semantically structured record of the actives, routes and tours of users, including actual activities and those recommend by the system and other users. The initial dataset will be created by NG staff, the NG CROSSCULT research fellow, and other CROSSCULT researchers. The rest of the dataset will be created by the users and the CROSSCULT systems.</p> <p>The dataset will take the form of structured XML files. The format will allow reuse and further data enrichment. The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing.</p> <p>The dataset from users recommending things for other users /curating their own tours and activities will be anonymised, unless the original user has chosen to publish their details against the recommended activities. The storage will be organised by assigning an anonymised identifier to every participant.</p>
<i>Data sharing</i>	<p>All the data belonging to this dataset will be anonymised data, unless the information has already been specifically made public by the user(s). They will have no privacy concerns and can be made available to any interested researcher.</p> <p>Since full use of this dataset will be in combination with the <i>CROSSCULT-DS-Venue_Infrastructure_Collections</i> dataset, so there may be some security restrictions on its use that will need to be taken into account.</p> <p>Where possible, generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible and will be deposited in the project repositories to facilitate reuse.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be deposited in the project repositories to facilitate reuse.</p> <p>The NG already conducts backups every night to tape on site in a server room facility and the backup is also stored offsite to an external provider. Some of the costs for this will be covered by CROSSCULT during the project. At the end of the project final versions of the dataset will be archived as part of the ongoing NG backup procedure.</p>

Table 10. Pilot 1 multimedia images

CROSSCULT-DS-pilot_1_Multimedia_Images	Pilot 1 multimedia images
<i>Dataset description</i>	<p>This dataset will contain the multimedia resources, images provided by the venue in pilot 1 that will be used to create stories to be presented to the users. The dataset will consist of:</p> <ul style="list-style-type: none"> • A full set of 800 pixel collection images of National Gallery acquisitioned paintings (~2300). • Possible selective inclusion of technical images, details, samples (that need to be related to interpreted details). <p>All images provided to the system will be generated by an IIPImage/IIIF end-point. This will ensure that the most recent and appropriate images are always used by the CROSSCULT systems. This relates to new acquisitions, cleaned paintings and or just improved photography.</p> <p>This system could provide zoomable images or just simple static thumbnails and jpegs.</p> <p>Additional image resources could be created if necessary for the project by direct capture or recordings of the venue's assets. Additional video and alternative surface texture based multimedia may also become available, but these will be subject to specific additional agreements that would be organised by the NG.</p> <p>As the multimedia contents are subject to access restrictions, derived from the NG Intellectual Property policies, access will be granted wherever possible to allow worldwide researchers detailed analysis for the whole collection rather than individual paintings.</p> <p>Open resources and the attached metadata may be used in project publications if deemed appropriate.</p>
<i>Standards and metadata</i>	<p>The data will use the most well-known formats regarding multimedia resources. The IIIF and MPEG-7 standards will be used to describe the relevant details of the involved files, describing their technical characteristics and the more appropriate descriptions of their contents.</p> <p>This dataset will have been created by NG staff, as part of other related NG projects, the NG CROSSCULT research fellow, along with assistants from other CROSSCULT researchers.</p> <p>This dataset will be created as part of the CROSSCULT project along with other related Projects and an ongoing program within in the NG to develop and enhance its collection information.</p>
<i>Data sharing</i>	<p>The data of this dataset have different access levels at the time of writing.</p> <p>The full set of 800 pixel collection images is made available under a Creative Commons (CC) BY-NC-ND 4.0 license. Specific agreement/licence would need to be defined for the high-resolution images presented on the NG website.</p> <p>Further generated data and metadata needed to validate results in scientific publications and underpin published research findings as well</p>

	<p>as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be deposited in a data repository to be findable and accessible.</p> <p>Access to any augmentation of the dataset created outside the CROSSCULT APP will be as open as possible but specific agreements will be required.</p>
<i>Archiving and preservation (including storage and backup)</i>	All of the images and multimedia items included within CROSSCULT will also be part of the NG internal multimedia archive and will continue to be archived and made available, where appropriate for reuse within future projects related to the NG.

Table 11. Pilot 1 user tracking and observations

CROSSCULT-DS-pilot_1_User_Tracking_Observation	
<i>Dataset description</i>	<p>This dataset will contain all the data collected from observation and tracking location of the users participating in the experiments using the CROSSCULT app. It will include information about the users' interaction with the NG collection information. The data will be directly related to the actual user route information included in the <i>CROSSCULT-DS-pilot_1_NG_Routes_Tours</i> dataset.</p> <p>The dataset will comprise anonymised information about the users who will participate in the NG visits, using the CROSSCULT app. Additionally, for users who give their permission, information can be automatically extracted/retrieved from their devices location. User agreement must be obtained before tracking and processing data about the user.</p> <p>The dataset generated from the observation will provide information on user behaviour. The exact make-up of the fields included in this dataset will be determined as part of the work carried out within CROSSCULT.</p> <p>The dataset generated from the tracking location will provide information on the location of the user or proximity to a predefined location, the history of where a user has been and the time they spend at any location, how long users spend in a room, which rooms are visited more often (room's visitor traffic), etc. The data captured will be organised into hierarchical actual user events; Visits to the NG, visits to specific rooms, visits to specific paintings. This will be combined with details of any recorded digital activities.</p> <p>This dataset could be used by worldwide researchers to run experiments with similar or different profiles in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc.</p> <p>This dataset may underpin research publications derived from the project, as the characteristics of the users are determinant when evaluating the conclusions of the experience.</p>
<i>Standards and</i>	The locations of users will need to be recorded and dynamically linked to

<i>metadata</i>	<p>systems storing current paintings' locations. User-defined old locations will also need to be stored to allow users to explore alternative, virtual, presentations of the collection.</p> <p>The dataset will take the form of text files containing all the relevant information. This dataset will then take the form of structured XML files.</p>
<i>Data sharing</i>	<p>All the data belonging to this dataset will be anonymised data, unless the information has already been specifically made public by the user(s). They will have no privacy concerns and can be made available to any interested researcher.</p> <p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p> <p>Generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be deposited in the recommended project repository to facilitate reuse. The NG already conducts backups every night to tape on site in a server room facility and the backup is also stored offsite to an external provider.</p>

Table 12. Pilot 1 game scenarios

CROSSCULT-DS-pilot_1_Game_Scenarios	Pilot 1 game scenarios
<i>Dataset description</i>	<p>This dataset will contain information about the content available in the game scenarios for the users participating in the experiments of pilot 1 using the CROSSCULT app. It will have generic structures for quizzes in addition to a fixed set of questions, as a predefined set of questions, with default question types.</p> <p>The fixed set of questions will be selected in relation to the pilot 1 reflection points and the available related data. While users interact with the CROSSCULT APP, new sets of question will be generated based on user options, preferences, and activities and profiling. The dataset will also include anonymised data related to which quizzes are used and how well different users have performed in each attempted quiz.</p> <p>This completed dataset could be used by worldwide researchers to investigate gaming within museums down to specific such as time, user interest, and user's engagement with a cultural collection. Personalisation and context-awareness are also two important elements implemented in CROSSCULT technologies. It would also be possible to relate the results of quizzes to the user activities with the CROSSCULT app and the National Gallery, allowing researchers to explore how well different interactions with the collection relate to a user's ability to answer questions.</p>

	<p>This dataset may underpin research publications derived from the project addressing the user needs in places and contexts related to learning about history and culture.</p>
<i>Standards and metadata</i>	<p>This dataset will include details of the questions and answers provided to and by the users within the quizzes presented as part of pilot 1. The initial dataset will be created by NG staff, the NG CROSSCULT research fellow, along with assistants from other CROSSCULT researchers. The rest of the dataset will be created by the users and the CROSSCULT systems itself.</p> <p>The dataset will be presented on the form of XML files structuring the answers from every user. It will be organised by assigning an anonymised identifier to every participant.</p>
<i>Data sharing</i>	<p>All the data belonging to this dataset will be anonymised data, unless the information has already been specifically made public by the user(s). They will have no privacy concerns and can be made available to any interested researcher.</p> <p>The format of these files (XML) will allow reuse and further data enrichment. The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing.</p> <p>Generated data and metadata needed to validate results in scientific publications and underpin published research findings as well as other curated data and metadata that may be required for validation purposes and has longer-term value for reuse, will be made fully accessible.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be deposited in the recommended project repository to facilitate reuse. The NG already conducts backups every night to tape on site in a server room facility and the backup is also stored offsite to an external provider.</p>

5. Pilot 2 datasets

Table 13. Pilot 2 ontology extensions and literals

CROSSCULT-DS-pilot_2_Ontology_Extensions	Pilot 2 ontology extensions and literals
CROSSCULT-DS-pilot_2_Literals	
<i>Dataset description</i>	<p>These datasets will be used to model the knowledge about the items from the four venues participating in pilot 2. The datasets will be composed of the extensions of the upper-level ontology used to describe the types and relationships required and the actual literal data related to the collection.</p> <p>The basic data will be augmented by related concepts such as geographical data and related vocabulary terms and connected to external sources of information (Wikidata, Wikipedia) and other museums APIs (BM, Europeana).</p> <p>The ontology may be used for further semantic related projects to help define and describe similar collection and as part of more general semantic research. The literals may be used in further research to examine the relationships captured in the dataset, further explore the collections connections to the broader world history, further explore the visualisation of complex cultural heritage datasets and or as an example for how other collections might be described.</p>
<i>Standards and metadata</i>	<p>These datasets will be composed of the ontology(ies) used to describe the types and relationships required to capture the items from the venues participating in pilot 2.</p> <p>The final literal dataset will take the form of structured XML files, with the information modelled in OWL. The types and relationships captured within these files will be documented within the overall ontology, which will also take the form of structured XML. The overall ontology will be based on the CIDOC-CRM cultural heritage standard, along with other related ontologies and enhancements that have been selected and or developed within the CROSSCULT project.</p>
<i>Data sharing</i>	<p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The ontology extensions will be shared with any interested researcher through the project's general repository. The literals may be subject to access restrictions, derived from the venue's Intellectual Property policies. Open, unrestricted access will be granted wherever possible.</p>

Table 14. Pilot 2 visitor profiles

CROSSCULT-DS-pilot_2_Visitor_Profiles	Pilot 2 visitor profiles
<i>Dataset description</i>	<p>This dataset will include all the information used to create the profiles of the visitors participating in the experiments of pilot 2. It will comprise anonymised information about the visitors who participate in the games, including psycho-demographic data, cognitive style, estimations of historical knowledge and interests mined from social networks. Some information will be gathered through questionnaires and tests before playing the game. Additionally, for visitors who give their permission, the app will automatically extract information from social networks to infer interesting profile traits.</p> <p>The archived information will consist of XML files containing the answers to the questionnaires and the findings obtained through social mining. The actual information posted in the social network accounts will never be stored in order to prevent any identity guessing.</p> <p>This dataset could be used by worldwide researchers to run experiments with similar or different profiles in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc. It is expected that the dataset will contain up to a few hundred profiles. All the information will be anonymised.</p> <p>This dataset may underpin research publications derived from the project, as the characteristics of the users are determinant when evaluating the conclusions of the experience.</p>
<i>Standards and metadata</i>	<p>The information of this dataset will be provided by the participants through questionnaires and through direct access to the information of their social networks. The stored information will consist of XML files structuring the answers from every user and the data mined from their social networks. When appropriate, the answers could be normalised through well-known references like the Likert scale.</p> <p>The storage will be organised by assigning an anonymised identifier to every participant. The naming scheme for the individual XML files will not take into account any version information, because the profiles are not expected to evolve in pilot 2. Nonetheless, the whole collection of the files will follow the naming scheme specified in the “<i>Naming conventions and document versioning</i>” section 4.2.2.3 of CROSSCULT’s Deliverable D1.1 (“<i>Quality and risk management plan</i>”).</p>
<i>Data sharing</i>	<p>All the data belonging to the dataset will be anonymised data. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access data online with no restriction at all. The XML files may be opened with any text editor or more advanced tools.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be stored and made available to worldwide researchers through the project repositories (see Section 2). The volume of the information will be small,</p>

	<p>up to several Kbytes for each participant at the most.</p> <p>Since the repositories are free, no costs are foreseen for long-term storage of this dataset.</p>
--	--

Table 15. Pilot 2 multimedia contents

CROSSCULT-DS-pilot_2_Multimedia_Contents	Pilot 2 multimedia contents
<i>Dataset description</i>	<p>This dataset includes the multimedia resources linked to the venues involved in pilot 2 that may be used to create stories to be told to the participants in the experiments. The intended resources include text, images, video and audio clips, animations, 3D shapes, AR contents, external links, and any other formats. Descriptive metadata will contain semantic characterizations for every resource.</p> <p>The multimedia resources will use the most common formats (gif, jpeg, png, avi, mpg, mp4, doc, txt, odt, ...), so integration and reuse difficulties are not foreseen. Transcoding may be used wherever necessary. The descriptive metadata will rely on MPEG-7 standard vocabularies as far as possible.</p> <p>Some of the resources will be provided by the venues themselves. Additional material will be linked from open online sites like Wikipedia. It is foreseen to include short clips (up to 30 seconds long) extracted from classical European movies.</p> <p>Both the multimedia contents and the descriptive metadata may be subject to access restrictions, derived from the venues' Intellectual Property policies. Open, unrestricted access will be granted wherever possible to allow worldwide researchers to arrange experiments with the same or different combinations of materials in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc.</p> <p>Open resources and the attached metadata may be used in project publications if deemed appropriate.</p>
<i>Standards and metadata</i>	<p>Most of the involved resources are expected to belong to the digital archives of the participant venues. In any case, additional resources could be created if necessary for the project by direct capture or recordings of the venues assets (sculptures, paintings, vessels, etc.). The stored data will use the most common formats for multimedia content. If necessary, appropriate transcoding will be applied to the available archives.</p> <p>The MPEG-7 standard will be used to describe the relevant details of the involved files, describing their technical characteristics and the more appropriate descriptions of their contents. The original metadata from venues databases that describe the resources will be transformed and adapted to the MPEG-7 standard. In addition, a global metadata file will be created that will declare all the items contained in the dataset, their</p>

	<p>structure, ordering scheme and access terms.</p> <p>The files will be organised by folders according to the participant venues, and, innermost, in structured folders according to the type of asset. For each file corresponding to a multimedia resource there will exist a metadata archive containing its description.</p>
<i>Data sharing</i>	<p>Access to the dataset will be possible through the global asset declaration provided in the root of the repository that will announce all the items of the dataset, their name, classification, associated metadata and access rights.</p> <p>The data will be shared with any interested researcher through the project's general repository. Both the multimedia contents and the descriptive metadata may be subject to access restrictions, derived from the venues' Intellectual Property policies. Open, unrestricted access will be granted wherever possible.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The information volume that will be stored in this dataset is difficult to foresee at this stage. This will be strongly dependant on the final scripts of the experimental pilots and the resources selected from the venues repositories and other sources. The estimation is around several Gbytes. Since the project repositories are free, no costs are foreseen for long-term storage of this dataset.</p>

Table 16. Pilot 2 game scenarios

CROSSCULT-DS-pilot_2_Game_Scenarios	Pilot 2 game scenarios
<i>Dataset description</i>	<p>This dataset will contain information about the content available in the game scenarios for the users participating in the experiments of pilot 2. It will contain predefined graphs of concepts and relationships to explore, attached sets of choices for the concepts that will be left blank (including different sets for different individual/team profiles and locations) and attached multimedia contents. The scenarios will be developed by CROSSCULT humanities experts.</p> <p>This dataset may be used by worldwide researchers, e.g. as a reference to evaluate the operation of algorithms trying to automate any part of the experts' design work. As such, this dataset may underpin research publications derived from the project addressing the user needs in places and contexts related to learning about history and culture.</p>
<i>Standards and metadata</i>	<p>This dataset will include details of the questions and answers provided to and by the users within the quizzes presented as part of pilot 2. The scenarios will be developed by CROSSCULT humanities experts.</p> <p>The dataset will be presented on the form of XML files indicating the venues involved, the structuring the concepts and relationships in the graph, the concepts left blank and the possible sets of answers.</p>
<i>Data sharing</i>	<p>The design of the game scenarios will be freely available to interested researchers. The format of the files (XML) will allow reuse and further</p>

	data enrichment.
<i>Archiving and preservation (including storage and backup)</i>	All the information files belonging to this dataset will be deposited in the recommended project repository to facilitate reuse.

Table 17. Pilot 2 visitor interactions and staff observations

CROSSCULT-DS-pilot_2_Visitor_Interactions_Staff_Observations	Pilot 2 visitor interactions and staff observations
<i>Dataset description</i>	<p>This dataset will include the contributions of participants during the experiments. The information (linked to the corresponding profiles from the CROSSCULT-DS-pilot_2_Visitor_Profiles dataset) will comprise records of individual/collective interactions, plus observations from the guiding staff of the venues (e.g. about the mood of the team discussions, amount of idle time, reception of the micro-augmentations, etc.).</p> <p>The format of the data will be XML files structuring the different kind of contributions and contributors in every stage of the user experience.</p> <p>This dataset could be used by worldwide researchers to compare their own experiment results with the ones of the project and to extract conclusions in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc. Since all the information will be anonymised, privacy concerns will not be an issue.</p> <p>This dataset will be extremely important in any research publication derived from the project, as the data items yield indicators and measurements of the quality of the user experience. No additional information is foreseen, as this information cannot be expanded or enhanced by any external contribution.</p>
<i>Standards and metadata</i>	<p>The information of this dataset will be provided by the participants in the proposed experiences. The stored information will consist of XML files structuring all such contributions in the different stages of the experiment. The storage will be organised assigning an anonymised identifier to every participant, and grouping the different contributions by stage.</p> <p>There will be one file for each stage, grouping the different contributions of every participant. They will be differentiated by their role: visitor, expert, staff... The naming scheme will not take into account any version information, as this information is not expected to evolve.</p>
<i>Data sharing</i>	All the data belonging to the dataset will correspond to anonymised data. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access data online with no restrictions at all, accessing the repository without requesting any kind of previous permission.
<i>Archiving and</i>	The information volume that will be stored is quite small, several Kbytes

<i>preservation (including storage and backup)</i>	for each participant at the most, yielding up to a few Mbytes per experimentation session. Since the project repositories are free, no costs are foreseen for long-term storage of this dataset.
--	--

Table 18. Pilot 2 interview transcripts

CROSSCULT-DS-pilot_2_Interview_Transcripts	Pilot 2 interview transcripts
<i>Dataset description</i>	<p>This dataset provides the user responses to the interviews that will be performed during the evaluation phase of the project and will be designed around the unified theory of users acceptance of technology. Data will be collected during pilot 2 evaluations, consisting of anonymised transcripts of interviews.</p> <p>The dataset could be used by any interested researchers. It is expected that the dataset can contribute to an understanding of the users' acceptance of the CROSSCULT technology and to an article based on this study. It will also inform the evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin core set of metadata will be used to describe the dataset. It will be created during the pilot 2 tests in order to understand how users accept and perceive the concept. It will be made available as a set of .docx documents for each transcription.
<i>Data sharing</i>	The dataset will be made available on the general project repository with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	The files will be available to worldwide researchers through the project repositories (see Section 2). No costs are foreseen for long-term storage.

6. Pilot 3 datasets

Table 19. Pilot 3 ontology extensions and literals

CROSSCULT-DS-pilot_3_Ontology_Extensions	Pilot 3 ontology extensions and literals
CROSSCULT-DS-pilot_3_Literals	
<i>Dataset description</i>	<p>These datasets will be used to model the knowledge about the items from the Archaeological Museum of Tripolis, that will be used in pilot 3. The datasets will be composed of the extensions of the upper-level ontology used to describe the types and relationships required and the actual literal data related to the collection.</p> <p>The basic data will be augmented by related concepts such as geographical data and related vocabulary terms and connected to external sources of information (Wikidata, Wikipedia) and other museums APIs (BM, Europeana).</p> <p>The ontology may be used for further semantic related projects to help define and describe similar collection and as part of more general semantic research. The literals may be used in further research to examine the relationships captured in the dataset, further explore the collections connections to the broader world history, further explore the visualisation of complex cultural heritage datasets and or as an example for how other collections might be described.</p>
<i>Standards and metadata</i>	<p>These datasets will be composed of the ontology(ies) used to describe the types and relationships required to capture the items from the venues participating in pilot 3.</p> <p>The final literal dataset will take the form of structured XML files, with the information modelled in OWL. The types and relationships captured within these files will be documented within the overall ontology, which will also take the form of structured XML. The overall ontology will be based on the CIDOC-CRM cultural heritage standard, along with other related ontologies and enhancements that have been selected and or developed within the CROSSCULT project.</p>
<i>Data sharing</i>	<p>The XML files may be opened with any text editor or more advanced tools (Oxygen) for further processing. The RDF triples will allow reuse and further data enrichment. These files may be opened with a web browser and freely-available tools.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>The ontology extensions will be shared with any interested researcher through the project's repository. The literals may be subject to access restrictions, derived from the venue's Intellectual Property policies. Open, unrestricted access will be granted wherever possible.</p>

Table 20. Pilot 3 visitor profiles

CROSSCULT-DS-pilot_3_Visitor_Profiles	Pilot 3 visitor profiles
<i>Dataset description</i>	<p>This dataset will include all the information used to create the profiles of the visitors participating in the experiments of pilot 3. It will comprise anonymised information about the visitors who participate in the museum visits, including psycho-demographic data, cognitive style, and interests mined from social networks. Part of this information will be gathered through games and quizzes before the user experiences. Additionally, for visitors who give their permission, the app will automatically extract information from social networks to infer interesting profile traits.</p> <p>The archived information will consist of XML files containing the answers to the quizzes and games and the findings obtained through social mining. The actual information posted in the social network accounts will never be stored in order to prevent any identity guessing.</p> <p>This dataset could be used by worldwide researchers to run experiments with similar or different profiles in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc. It is expected that the dataset will contain up to a few hundred profiles. Since all the information will be anonymised, privacy concerns will not be an issue.</p> <p>This dataset may underpin research publications derived from the project, as the characteristics of the users are determinant when evaluating the conclusions of the experience.</p>
<i>Standards and metadata</i>	<p>The information of this dataset will be provided by the participants through quizzes and games and through direct access to the information of their social networks.</p> <p>The stored information will consist of XML files structuring the answers from every user and the data mined from their social networks. The storage will be organised by assigning an anonymised identifier to every participant. The naming scheme for the individual XML files will not take into account any version information, because the profiles are not expected to evolve in pilot 3. Nonetheless, the whole collection of files will follow the naming scheme specified in the “<i>Naming conventions and document versioning</i>” section 4.2.2.3 of CROSSCULT’s Deliverable D1.1 (“<i>Quality and risk management plan</i>”).</p>
<i>Data sharing</i>	<p>All the data belonging to the dataset will be anonymised data. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access data online with no restriction at all. The XML files may be opened with any text editor or more advanced tools.</p>
<i>Archiving and preservation (including storage and backup)</i>	<p>All the information files belonging to this dataset will be stored and made available to worldwide researchers through the project repositories (see Section 2). The volume of the information will be small,</p>

	<p>up to several Kbytes for each participant at the most.</p> <p>Since the repositories are free, no costs are foreseen for long-term storage of this dataset.</p>
--	--

Table 21. Pilot 3 multimedia contents

CROSSCULT-DS-pilot_3_Multimedia_Contents	Pilot 3 multimedia contents
<i>Dataset description</i>	<p>This dataset includes the multimedia resources linked to the venues involved in pilot 3 that may be used to create stories to be told to the participants in the experiments. The resources include text, images, video and audio clips, animations, 3D shapes, AR contents, external links, and any other formats. The multimedia resources will use the most common formats (gif, jpeg, png, avi, mpg, mp4, doc, txt, odt, ...), so integration and reuse difficulties are not foreseen. Transcoding may be used wherever necessary.</p> <p>Descriptive metadata (relying on MPEG-7 standard vocabularies as far as possible) will contain semantic characterizations for every resource.</p> <p>Some of the resources will be provided by the venue itself. Additional material will be linked from open online sites like Wikipedia, Wikimedia, etc. It is foreseen to include short clips (always respecting copyright laws) from Greek music, ancient to traditional and modern, as well as world music.</p> <p>Both the multimedia contents and the descriptive metadata may be subject to access restrictions, derived from the venues' Intellectual Property policies. Open, unrestricted access will be granted wherever possible to allow worldwide researchers to arrange experiments with the same or different combinations of materials in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc. Open resources and the attached metadata may be used in project publications if deemed appropriate.</p>
<i>Standards and metadata</i>	<p>Most of the involved resources are expected to belong to the digital archives of the participant venues. In any case, additional resources could be created if necessary for the project by direct capture or recordings of the venues assets (sculptures, paintings, vessels, etc.). The stored data will use the most common formats for multimedia content. If necessary, appropriate transcoding will be applied to the available archives.</p> <p>The MPEG-7 standard will be used to describe the relevant details of the involved files, describing their technical characteristics and the more appropriate descriptions of their contents. The original metadata from venues databases that describe the resources will be transformed and adapted to the MPEG-7 standard. In addition, a global metadata file will be created that will declare all the items contained in the dataset, their structure, ordering scheme and access terms.</p>

	The files will be organised by folders according to the participant venues, and, innermost, in structured folders according to the type of asset. For each file corresponding to a multimedia resource there will exist a metadata archive containing its description.
<i>Data sharing</i>	<p>Access to the dataset will be possible through the global asset declaration provided in the root of the repository that will announce all the items of the dataset, their name, classification, associated metadata and access rights.</p> <p>The data will be shared with any interested researcher through the project's general repository. Both the multimedia contents and the descriptive metadata may be subject to access restrictions, derived from the venues' Intellectual Property policies. Open, unrestricted access will be granted wherever possible.</p>
<i>Archiving and preservation (including storage and backup)</i>	The information volume that will be stored in this dataset is difficult to foresee at this stage. This will be strongly dependant on the final scripts of the experimental pilots and the resources selected from the venues repositories and other sources. The estimation is around several Gbytes. Since the project repositories are free, no costs are foreseen for long-term storage of this dataset.

Table 22. Pilot 3 visitor interactions and staff observations

CROSSCULT-DS-pilot_3_Visitor_Interactions_Staff_Observations	Pilot 3 visitor interactions and staff observations
<i>Dataset description</i>	<p>This dataset will include the contributions of participants during the experiments. The information (linked to the corresponding profiles from the <i>CROSSCULT-DS-pilot_3_Visitor_Profiles</i> dataset) will comprise records of individual/collective interactions during the stages of the experiment, plus observations from the guiding staff of the venues (e.g. about the mood of the team discussions, amount of idle time, reception of the micro-augmentations, etc.).</p> <p>The form of the data will be XML files structuring the different kind of contributions and contributors in every stage of the experience.</p> <p>This dataset could be used by worldwide researchers to compare their own experiment results with the ones of the project and to extract conclusions in order to check the learning differences according to distinct parameters: local culture, previous background, relationship to the historical contents of the venues, etc. Since all the information will be anonymised, privacy concerns will not be an issue.</p> <p>This dataset will be extremely important in any research publication derived from the project, as the data items yield indicators and measurements of the quality of the experience. No additional information is foreseen, as this information cannot be expanded or enhanced by any external contribution.</p>
<i>Standards and</i>	The information of this dataset will be provided by the participants in

<i>metadata</i>	<p>the proposed experiences. The stored information will consist of XML files structuring all such contributions in the different stages of the experiment. The storage will be organised assigning an anonymised identifier to every participant, and grouping the different contributions by stage.</p> <p>There will be one file for each stage, grouping the different contributions of every participant. They will be differentiated by their role: visitor, expert, staff... The naming scheme will not take into account any version information, as this information is not expected to evolve.</p>
<i>Data sharing</i>	All the data belonging to the dataset will correspond to anonymised data. They will have no privacy concerns and can be made available to any interested researcher. Anyone will be able to access data online with no restriction at all, accessing to the repository without requesting any kind of previous permission.
<i>Archiving and preservation (including storage and backup)</i>	The information volume that will be stored is quite small, several Kbytes for each participant at the most, yielding up to a few Mbytes per experimentation session. Since the project repositories are free, no costs are foreseen for long-term storage of this dataset.

Table 23. Pilot 3 interview transcripts

CROSSCULT-DS-pilot_3_Interview_Transcripts	Pilot 3 interview transcripts
<i>Dataset description</i>	<p>This dataset provides the user responses to the interviews that will be performed during the evaluation phase of the project and will be designed around the unified theory of users acceptance of technology. Data will be collected during pilot 3 evaluations, consisting of anonymised transcripts of interviews.</p> <p>The dataset could be used by any interested researchers. It is expected that the dataset can contribute to an understanding of the users' acceptance of the CROSSCULT technology and to an article based on this study. It will also inform the evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin core set of metadata will be used to describe the dataset. It will be created during the pilot 3 tests in order to understand how users accept and perceive the concept. It will be made available as a set of .docx documents for each transcription.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	The files will be available to worldwide researchers through the project repositories (see Section 2). No costs are foreseen for long-term storage.

7. Pilot 4 datasets

Table 24. Pilot 4 ontology extensions and literals

CrossCult-DS-pilot_4_ontology_extensions	Pilot 4 ontology extensions and literals
CrossCult-DS-pilot_4_literals	
<i>Dataset description</i>	This dataset will be used to model knowledge about the multimedia items used in pilot 4. This dataset will be composed of upper-level ontology classes and a set of specialisation (extensions) required for describing the types and relationships required and the actual literal data related to the collection.
<i>Standards and metadata</i>	These datasets will be composed of the ontologies used to describe the types and relationships required to capture the items from the multimedia resources participating in pilot 4.
<i>Data sharing</i>	The ontology extensions will be shared with any interested researcher through the project's general repository. The literals may be subject to access restrictions. Open, unrestricted access will be granted wherever possible.
<i>Archiving and preservation (including storage and backup)</i>	This dataset will be deposited in the recommended project repository to facilitate reuse (see section 2).

Table 25. Pilot 4 geolocated object discovery sequences

CROSSCULT-DS-pilot_4_Sequence_Objects	Pilot 4 geolocated object discovery sequences
<i>Dataset description</i>	<p>This dataset identifies the sequence in which objects are discovered in the built environment and their location in the city. It keeps track of the order and locations in which of objects were discovered by each player and provides and aggregate of the results. The data will be collected during pilot 4 evaluations, and consist of a geolocated matrix based on individual long-lat locations (WGS84) for Malta and Valetta.</p> <p>The dataset could be used by geographers, Ville de Luxembourg, spatial planners and even tourist guides or people with tourist activities in town (where do people go and what is the relation to place). The dataset will contribute to an article based on the relevance of place for reflective history practices. It will also inform the CROSSCULT evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin core set of metadata will be used to describe the dataset. It will be created during the pilot 4 testing as users interact with the app and play the game. Then, it will be made available as a .xls file

	with coordinate data in WGS84 format.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	For the case of Luxembourg it it makes sense then the data will be uploaded and to made available on data.public.lu otherwise they will be deposited in the recommended project repository to facilitate reuse.

Table 26. Pilot 4 reflections on History

CROSSCULT-DS-pilot_4_Reflections_History	Pilot 4 reflections on History
<i>Dataset description</i>	<p>This dataset aggregates the personal historic reflections that are collected and curated as players interact with pilot 4 app. Pilot 4 will use open questions to engage users in reflection and (re)interpretation as they discover historic objects hidden in the built environment. During the evaluation period of pilot 4 the responses to the questions will be stored in the project database. The data will be collected during pilot 4 evaluations. It will contain anonymised responses containing fields such as time, date, original object and question and associated responses, location response was provided and the city.</p> <p>The dataset could be used by public historians, social scientists and geographers, the Ville de Luxembourg, spatial planners and even tourist guides or people with tourist activities in town (where do people go and what is the relation to place). It is expected that the dataset will contribute to an article and conference paper based on using serious geogames for engaging in reflective history practices. It will also inform the CROSSCULT evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin Core set of metadata will be used to describe the dataset. It will be created during the pilot 4 testing as users interact with the app and play the game. Then, it will be made available as a .xls or .XML file.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	For the case of Luxembourg if it makes sense then the data will be uploaded and made available on data.public.lu otherwise they will be deposited in the recommended project repository to facilitate reuse.

Table 27. Pilot 4 collective memory stories

CROSSCULT-DS-pilot_4_Collective_Memory_Stories	Pilot 4 collective memory stories
<i>Dataset description</i>	This dataset will provides the user-generated content that is produced by pilot 4's users when they encounter a story treasure card. The data will be collected during pilot 4 evaluations, and will take the form of a

	<p>geolocated matrix based on individual stories with long-lat locations (WGS84) for Malta and Valetta.</p> <p>The dataset could be used by the general public, teachers and educators, museums and organisations concerned with public history and or migration as well as research in public historians, social scientists and geographers, the Ville de Luxembourg, spatial planners and even tourist guides or people with tourist activities in town (where do people go and what is the relation to place).</p> <p>It is expected that the dataset will contribute to an article based on relevance of place for reflective history practices. It will also inform the evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin Core set of metadata will be used to describe the dataset. It will be created during the pilot 4 testing as users interact with the app and play the game. Then, it will be made available as a .xls or.XML, geolocated data will be assigned coordinate data in WGS84 format.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	For the case of Luxembourg if it makes sense then the data will be uploaded and made available on data.public.lu otherwise they will be deposited in the recommended project repository to facilitate reuse.

Table 28. Pilot 4 questionnaire responses

CROSSCULT-DS-pilot_4_ Questionnaire_ Responses	Pilot 4 questionnaire responses
<i>Dataset description</i>	<p>This dataset provides the user responses to the evaluation framework questionnaire that will be collated during the evaluation phase of the project and will be designed around the unified theory of acceptance and use of technology.</p> <p>Data will be collected during pilot 4 evaluations, consisting of an anonymised matrix of responses containing variables related to social demographics, usability of the technology, technology habits of the users, perceptions of trust of the contents, user experience...</p> <p>The dataset could be used by teachers, public historians, social scientists and geographers, the Ville de Luxembourg, spatial planners and even tourist guides or people with tourist activities in town (where do people go and what is the relation to place).</p> <p>It is expected that the dataset can contribute to an understanding of the users' acceptance of the CROSSCULT technology and to an article based on this study. It will also inform the evaluation framework.</p>
<i>Standards and</i>	A simple Dublin Core set of metadata will be used to describe the dataset. It will be created during the pilot 4 tests in order to understand

<i>metadata</i>	how users accept the concept. It will be made available as a .xls or .XML.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	For the case of Luxembourg if it makes sense then the data will be uploaded and made available on data.public.lu otherwise they will be deposited in the recommended project repository to facilitate reuse.

Table 29. Pilot 4 interview transcripts

CROSSCULT-DS-pilot_4_Interview_Transcripts	Pilot 4 interview transcripts
<i>Dataset description</i>	<p>This dataset provides the user responses to the interviews that will be performed during the evaluation phase of the project and will be designed around the unified theory of users acceptance of technology. Data will be collected during pilot 4 evaluations, consisting of anonymised transcripts of interviews.</p> <p>The dataset could be used by both researchers and teachers including public historians, social scientists and geographers, the Ville de Luxembourg, spatial planners and even tourist guides or people with tourist activities in town (where do people go and what is the relation to place).</p> <p>It is expected that the dataset can contribute to an understanding of the users' acceptance of the CROSSCULT technology and to an article based on this study. It will also inform the evaluation framework.</p>
<i>Standards and metadata</i>	A simple Dublin Core set of metadata will be used to describe the dataset. It will be created during the pilot 4 tests in order to understand how users accept and perceive the concept. It will be made available as a set of .docx documents for each transcription.
<i>Data sharing</i>	The dataset will be made available on the general project repository made available with a Creative Commons with attribution license.
<i>Archiving and preservation (including storage and backup)</i>	For the case of Luxembourg it is possible that the data will be made available on data.public.lu otherwise they will be deposited in the recommended project repository to facilitate reuse.

8. DMP Updates

This Data Management Plan will be updated by M18 (August 2017) and M36 (February 2018), to account for any additions and/or changes not identified hitherto. The dataset descriptions will be completed and further homogenised together, and new ones will be added when needed.