



D3.4 Data Management Plan

INCULTUM Project 2021-2024 (Grant Agreement n. 101004552)

Data Management Coordinator

Maurizio Toscano (maurizio@eachtra.ie)

Data Analysis and Statistics Team

Karol Borowiecki (kjb@sam.sdu.dk)

Maja Uhre Pedersen (mup@sam.sdu.dk)

Sara Mitchell (sabe@sdu.dk)

Context

Partner responsible for the deliverable	EACHTRA
Deliverable author(s)	Maurizio Toscano
Deliverable version number	2.0
Dissemination Level	Public

History

Change log			
Version	Date	Author(s)	Reason for change
0.1	01/11/2021	Toscano M. and Mitchell S.	First draft
0.2	19/11/2021	VV. AA.	Info provided by INCULTUM partners
1.0	30/11/2021	Toscano M.	External review
1.1	04/12/2023	Toscano M.	Revision for final deliverable
1.2	31/01/2023	VV. AA.	Info provided by INCULTUM partners
2.0	21/04/2023	Toscano M.	External review and Zenodo Community

Release approval			
Version	Date	Name & organization	Role
1.0	29/11/2021	Borowiecki Karol - SDU	Data Analysis and Statistics WP Leader
2.0	22/04/2023	Borowiecki Karol - SDU	Data Analysis and Statistics WP Leader

Introduction	4
1. Data summary	5
2. FAIR data	7
2.1 Findability of data/research outputs	7
2.2 Accessibility of data/research outputs:	7
2.3 Interoperability of data/research outputs	8
2.4 Reusability of data/research outputs:	8
3. Allocation of resources	10
4. Data security	11
5. Ethical aspects	12
6. Open Access to scientific publications	14
7. Re-used datasets	15
8. Primary Datasets collected	16
9. Summary of Datasets not released as Open Data	18

Introduction

INCULTUM project participated in the Open Research Data Pilot (ORDP) to improve and maximise access and re-use of research data generated by the project and has taken into account the need to balance openness and protection of project's results, scientific data, commercialisation and IPR, privacy concerns and security, as well as data management and preservation questions.

The INCULTUM consortium strongly support Open Science and in the benefits that the European innovation ecosystem and economy can draw from allowing the reusing of data at a larger scale. Despite this, while carrying out pilot activities, individual partners have managed personal information and worked with local groups, collecting personal data, which rise privacy and ethical issues. In addition to that, some of the results are potentially exploitable to the benefit of the pilots, contributing to the outreach and impact of the project activities. Each case has been evaluated individually and has contributed to defining the INCULTUM project's data management policy. The purpose of this Data Management Plan (DMP) is to illustrate this data management policy and describe the results obtained in terms of datasets.

The principles stated in the first version of this DMP, published on 29/11/2021, guided data management and protection during the lifetime of the project, to be compliant with EU principles and standards, and with relevant national data protection laws and institutional data management policies, and to support the achievement of project's objectives.

The DMP has been so far a living document, which has evolved and has been regularly reviewed and updated during the lifespan of the project, particularly whenever significant milestones have been reached. This first version of the DMP included an overview of the datasets to be produced or collected by the project and the specific conditions that were foreseen for them. Most of what was planned, in terms of data collection, has been achieved. As foreseeable, some mismatch also occurred, and the project implementation demonstrated in some cases that a different approach in data collection was better suited to achieve pilots and general objectives. In several cases, that was the result of a positive contamination among pilots' solutions, achieved with regular cross-cutting onsite and online meetings among pilots.

An important decision, taken at the project level during a discussion carried out in an internal workshop on Open Access and Open Data, has been to centralise the publication of datasets, publications and other outputs in a dedicated Zenodo community (<https://zenodo.org/communities/incultum/>), instead of relying on separate repositories, under the responsibility of each partner.

1. Data summary

The INCULTUM project (Grant Agreement n. 101004552) has collected and managed datasets related to cultural tourism in the pilot sites. The general purpose of data collection has been to provide measurable econometric evidence of the ex-ante and ex-post pilots' situations and the impact of INCULTUM innovations, as well as to compare trans-border pilots. Data collection, in particular in relation to local residents and tourist perceptions of pilot areas, has been crucial in order to reach the planned objectives, for example: developing specific innovative pilot solutions based on cooperation and participatory approaches; fostering hidden potentials in remote, peripheral or deindustrialized areas; identify and compare relevant drivers and barriers that account for the success or failure of participatory models; make policy recommendations; implement data analysis and statistics, including socio-economic measures; implement a communication plan to reach out to the widest possible range of local actors.

Most of the datasets collected are composed of structured alphanumeric data in [XLS and CSV format](#) (21 records). The second most popular type of data collected are rural tracks, each one published both in [GPX and KLM format](#) (10 records). Many GPX files have also been collected by the Swedish pilot, for the purpose of tracking the trajectories of tourists in specific locations. Some of these datasets are also accompanied by PDFs, whose purpose is to explain the methodology and the tools used in data collection. As for the total volume of data generated, this metric is not a valuable element, as all datasets are very small in size (less than 5MB), while 1 single dataset is very large, with over 11GB of data collected from Tripadvisor, within WP3.

In total, the project collected over 50 datasets (not counting list of participants), which is well beyond the 30 datasets expected to be generated at the beginning of the project; 6 datasets (from previous projects or official statistics) were finally reused, in line with those expected in the first version of the DMP.

In terms of data utility, collected records served, and can keep serving, different purposes can be grouped as follows:

1. to contact local stakeholders for involvement in collaborative activities, field surveys and community meetings.
2. to gather local residents needs and their perception of cultural heritage in pilot areas.
3. to obtain feedback indicators of satisfaction from the tourists visiting the pilots and to measure the impact of implemented actions.

4. to improve the cultural tourism experience and propose new activities within the pilot areas.
5. for the creation of new culturally significant tourist routes and rural trails, improving the existing network of rural tracks.
6. to provide a stronger basis for evaluating, utilizing, and generalizing the results gained and innovations tested by the pilots.
7. to measure the evolution of the quantitative and qualitative parameters previously identified.
8. to understand tourism movements.
9. to obtain comparable international tourism data of peripheral areas and provide insights into peripheral tourism.

Possible beneficiaries of the published datasets are:

1. local residents and community groups interested in replicating or maintaining INCULTUM solutions.
2. tourists looking for alternative areas to visit, slow tourism and experiences involving local groups.
3. researchers who want comparable data across countries on tourism at a very disaggregated level.
4. researchers in the fields of tourism, place branding, and language comprehension and innovation.
5. tour operators, agencies and entrepreneurs in the tourism industry, in particular those operating at a local scale.
6. project's pilots: cross-fertilization among pilots has been sought during the project, but because most of the data arrived after the 2023 touristic season, it will continue beyond the project lifetime.
7. municipalities, regional entities, and policymakers.
8. schools, cultural institutions and non-profit organizations.
9. cultural and creative industries.

2. FAIR data

Data management in INCULTUM has been carried in line with FAIR data principles, which are guiding principles on how to make data Findable, Accessible, Interoperable and Reusable, formulated by [Force11](#) as a useful framework for thinking about sharing data in a way that will enable maximum use and reuse. The usage of Zenodo, which is a platform built and operated by CERN and OpenAIRE, as the common repository for all project's public output, helped us [ensuring compliance with FAIR principles](#).

2.1 Findability of data/research outputs

To ensure findability, we have assigned persistent IDs to publicly shared datasets ([DOI](#)), enriched them with metadata, and registered all of them in Zenodo, which is a reliable searchable resource for long-term archiving.

All datasets have been annotated with the corresponding metadata standards (the Zenodo's metadata used is compliant with [DataCite's Metadata Schema](#) minimum and recommended terms, with a few additional enrichments) in order to describe the dataset's context, license, condition & characteristics.

All researchers involved in the INCULTUM consortium have registered themselves at [ORCID](#) to obtain a personal persistent author identifier and this identifier has been used for all their data publications and datasets.

2.2 Accessibility of data/research outputs:

In order to ensure accessibility of data and research outputs, all published datasets are retrievable using the DOI standard protocol, with accessible metadata, assigned by Zenodo to [all datasets published](#).

In general, datasets produced have been released as Open Data, with few exceptions that are described in Chapter 9.

When specific proprietary software was used to manipulate data (ex. SPSS), the resulting datasets were exported in a standard open format, to improve accessibility. In those cases where it was not self-explanatory how to read the data, documentation in PDF has been provided, explaining the methods and conventions used, in order to facilitate data manipulation and usage.

2.3 Interoperability of data/research outputs

To ensure the interoperability of the underlying, collected and generated data, the project has relied on data standards to represent and describe the data. These standards have been discussed and agreed upon during an internal session linked to the [Data workshop carried out in March 2022](#), and then refined in the WP3 session carried out in the project's second global meeting in Baza, in June 2022. The chosen standards comply with the recommendations of the scientific community, adapted to the reality and the needs of INCULTUM pilots.

Compliance with ISO standard ISO 15489-1:2001 information and documentation -- records management, as well as directive 95/46/EC about the protection of personal data, have been sought.

During the project, the performance of data management has been monitored and deviations from the DMP and the agreed standards have been managed and solved.

Where relevant, a README file has been attached to the published dataset, in addition to the standard description in Zenodo, to help ensure that data can be correctly interpreted and re-analysed by others.

2.4 Reusability of data/research outputs:

Data produced by the project has been published according to *Creative Commons* licenses, encouraged by the European Commission in order to grant adequate and useful licensing solutions. Sharing and re-use of data is supported by the employment of Creative Commons licenses, especially [Attribution-NonCommercial-ShareAlike 4.0 International](#) (CC BY-NC-SA 4.0) and [Attribution 4.0 International](#) (CC BY 4.0). The [BY-NC-SA license](#) lets others remix, adapt, and build upon the work non-commercially, as long as they credit the author and license their new creations under identical terms and has been preferred when the local uptake of innovative approaches based on those data in INCULTUM pilot areas could be affected, like the Irish pilot case.

Reusability has been further fostered by incorporating INCULTUM datasets as Open Data in the Zenodo repository, which is a long-term solution to ensure re-usability without a specific limited timeframe.

Each beneficiary has examined, according to Art. 27 of the Grant Agreement, the possibility or need to adequately protect its results. Only five datasets have not been released as Open data, as indicated in Chapter 9.

Only fully anonymised data has been shared with other researchers, using interoperable formats. In all processing of personal data, the consortium has complied with the Data Protection principles that are set out in the EU GDPR.

All data have been internally reviewed for completeness, consistency, and plausibility.

3. Allocation of resources

The decision to set up and employ a centralised collection for datasets, publications, deliverables and other outputs of the project, and to choose the Zenodo repository as the storage and archiving solution has also been taken to optimise the allocation and usage of resources.

According to Zenodo [terms of usage](#), data upload and storage are free, so no specific resources have been allocated to make INCULTUM data FAIR and ensure their long-term preservation.

As for personal information collected, all data have been handled only by qualified researchers under strict confidentiality agreements, assuring that data access, data protection and privacy standards comply with national and European regulations. All subjects included in the research have signed an informed consent form. Signed consent forms are archived in paper format, in local or institutional archives, and are scheduled to be destroyed 60 days after the project ends.

INCULTUM requirements regarding costs for data management have been basic and standard storage solutions or web hosting services have covered the project needs and have been covered by partner direct costs for purchasing.

Data collection during fieldwork has been carried out by researchers and technicians employed by INCULTUM partners, avoiding outsourcing or subcontracting. All secondary existing datasets employed during the project are free-to-use.

Data management responsibilities have been distributed in the INCULTUM project as follows:

- Data Management Coordination: Maurizio Toscano, EACHTRA
- Data Analysis and Statistics: Karol Jan Borowiecki, SDU
- Ethical aspects: University of Granada Ethics Committee
- Data protection officer (DPO): José Antonio Castillo Parrilla (UGR)
- Zenodo community management: Maurizio Toscano (EACHTRA) and Teresa Bonet García (UGR)

4. Data security

Because of the diversity of datasets involved and the pilot cases approach, INCULTUM hasn't had a single unified backup policy but has relied, instead, on different solutions, depending on the location of the primary dataset during the project implementation.

As a general rule, a 3-2-1 backup strategy has been applied, with at least three (3) copies of each dataset, storing two (2) backup copies on different storage media, with one (1) of them located offsite.

For data hosted on web hosting, hourly backups for the past 24 hours, daily backups for the past 30 days, weekly backups for 1 year and monthly backups until the end of the project have been produced. Older backups' cleaning has been performed automatically by a cron job. Data on the web hosting will be kept online at least for one year after the project's conclusion.

This approach has preserved the project from possible data loss and enabled disaster recovery, although that was never performed. For permanent long-term storage, archiving and security of data, the project decided to rely, as said, on Zenodo.

Where data containing personal information, even pseudo-anonymized, have been transferred across partners, the dataset has been sent in a compressed folder protected by a password and encrypted using 7-zip software. The corresponding password has been sent using a different communication channel.

5. Ethical aspects

In terms of Ethics and Research Integrity, INCULTUM beneficiaries comply with instructions given in the art. 34 of the Grant Agreement and with ethics requirements set out in the WP8 and related deliverables. In addition to that, ethical and security issues in INCULTUM have been managed according to Universidad de Granada (UGR) [Good Practice Code for Research](#). The project has received ethical approval by UGR Ethics Committee, after rigorous internal review, as stated in deliverable D8.6.

The consortium is aware of and will comply with European and national legislation and fundamental ethical principles, including those reflected in the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights and its Supplementary Protocols. The activities of the INCULTUM project have been carried out in compliance with The European Code of Conduct for Research Integrity and the European General Data Protection Regulation, GDPR (GDPR EU 2016/679). This Law applies equally to electronic and non-electronic information; the intention is to ensure that personal data is managed in a way that is fair to the 'data subject' and that their interests are properly protected. Personal data is information that relates to a living individual – the data subject – which can be used to identify them either directly or indirectly in combination with other data. Photographs and images of all kinds in which a data subject can be identified are also regarded as personal data for the purposes of the Law. No disclosure or exposure of personal data to third parties has taken place without the written consent of the data subject. Personal data has only been used for the purpose for which it was originally collected. INCULTUM informed consent form and information sheet are available as deliverable D8.1.

INCULTUM project has carefully followed all the requirements for protecting personal data and the data generated within the project, in concordance with the data protection principles:

- A. Lawfulness, fairness, and transparency: personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject.
- B. Purpose limitation: personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.
- C. Data minimisation: personal data shall be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.
- D. Accuracy: personal data shall be accurate and, where necessary, kept up to date.

- E. Storage limitation: personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.
- F. Integrity and confidentiality: personal data shall be processed in a manner that ensures appropriate security of personal data, including protection against unauthorised or unlawful processing and accident loss, destruction, or damage, using appropriate technical or organisational measures.
- G. Accountability: the controller shall be able to demonstrate compliance with GDPR.

6. Open Access to scientific publications

Each partner has complied with the [Guidelines on Open Access to Scientific Publication and Open Research Data in Horizon 2020](#) and with the instructions stated in the arts. 29.2 and 29.3 of the Grant Agreement. The Steering Committee has been responsible for monitoring the proper dissemination and exploitation of results. All project publications have been released under “gold open access” regulation. That is also the case for the INCULTUM monograph, which is expected to be published during 2024 by Routledge.

A machine-readable electronic copy of each published version or final peer-reviewed manuscript accepted for publication has been [deposited in the Zenodo INCULTUM community](#). This step has been followed even in the case of 'gold' open-access publishing, to ensure long-term preservation of the article.

Datasets and related publications have been properly cross-linked to each other, when relevant, using the ad-hoc feature “Related works” provided by Zenodo.

For additional details on scientific publications and other outputs published, please refer to D2.3 Report on dissemination and communication.

7. Re-used datasets

Dataset name	Provider	Reference link
REACH Good Practices	H2020 REACH project	https://www.open-heritage.eu/best-practices
Archaeological data and irrigation channels from Sierra Nevada municipalities	MEMOLA project	https://memolaproject.eu/resources#filter=undefined&Format=.geo&Area=sierra-nevada
Swedish Population Mobile Phone Mobility data	Statistics Sweden	https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/mona--statistics-swedens-platform-for-access-to-microdata/
Irrigation channel in the provinces of Granada and Almería (Spain)	Regadío Histórico	https://regadiohistorico.es/espacios-de-regadio
Population and Social Conditions	Elstat (Hellenic Statistical Authority)	https://www.statistics.gr/en/statistics/pop
Labour Market (Employment Unemployment Entrepreneurship)	NATIONAL INSTITUTE OF LABOUR AND HUMAN RESOURCES (NILHR)	https://lmd.eiead.gr/ANNUAL-REPORT-2019/

8. Primary Datasets collected¹

Partner	Total datasets produced	Datasets released as OpenData
Pilot #1 UGR ²	<ul style="list-style-type: none"> - Perception surveys pre trails-development - Feedback surveys post trails-development - 6 GPS tracks - Geodatabase of irrigation channels 	<ul style="list-style-type: none"> - Perception surveys - Feedback surveys - 6 GPS tracks
Pilot #2 UALG	<ul style="list-style-type: none"> - Residents interviews - Questionnaire to visitors - Water sites survey 	<ul style="list-style-type: none"> - 1 Questionnaire to residents - 1 Questionnaire to visitors - Water sites survey
Pilot #3 UMB	<ul style="list-style-type: none"> - Shopping model questionnaire - Tourism-related statistics from localities included in the Mining Treasures platform 	<ul style="list-style-type: none"> - 1 Questionnaire to visitors - Tourism-related statistics
Pilot #4 GAL	<ul style="list-style-type: none"> - Marketing research results - 6 GPS tracks 	<ul style="list-style-type: none"> - Marketing research results - 6 GPS tracks
Pilot #5 UNIPI	<ul style="list-style-type: none"> - Perception survey to Museum visitors 2022 - Perception survey 2023 ed. - EN and IT 	<ul style="list-style-type: none"> - 2 Perception surveys - 1 Training course
Pilot #6 BI	<ul style="list-style-type: none"> - Network of rural tracks - Tourism Observatory survey 	<ul style="list-style-type: none"> - 2 individual trails - 2 Perception survey 2023-2024 - 1 Network or rural paths
Pilot #7 THM	<ul style="list-style-type: none"> - Cultural/natural Resources in Aaos Valley - Population, Social Conditions & Labour Market - Perception surveys - 2 GPS track 	<ul style="list-style-type: none"> - 2 GPS trail tracks - 1 questionnaire to visitors - 4 questionnaires to stakeholders
Pilot #8 CERPHAAL	<ul style="list-style-type: none"> - 1 GPS track - Visitors' questionnaire data - Archival and ethnoarchaeological study - Geodatabase in ArcGIS of infrastructure data 	<ul style="list-style-type: none"> - 1 GPS track - Visitors' questionnaire data
Pilot #9 EACHTRA	<ul style="list-style-type: none"> - 1 Online survey - 9 onsite surveys - Onsite POI QR metrics - Transcribed headstones database - Geophysical surveys 	<ul style="list-style-type: none"> - 1 online survey - 9 onsite surveys - Onsite POI QR metrics

¹ Contact lists of local stakeholders and participants in pilots' activities were collected by all pilots and cannot be published as open data because they contain personal information; for simplification, they have not been included in this table.

² The links provide the full list of records for each partner, including datasets, publications, deliverables, and other outputs. Unfortunately, Zenodo does not allow to filter by community, so few of the links also include some record not produced by INCULTUM.

<u>Pilot #10 UU</u>	<ul style="list-style-type: none"> - Surveys on perceptions of tourists' experiences - 3 GPS track bundles - Clicks from QR codes 	<ul style="list-style-type: none"> - 7 surveys of tourists - 3 GPS trajectories of tourists
<u>CBS</u>	<ul style="list-style-type: none"> - Campina de Faro tourists' survey 	<ul style="list-style-type: none"> - Campina de Faro tourists' survey
<u>SDU</u>	<ul style="list-style-type: none"> - Tripadvisor - Google Trends 	<ul style="list-style-type: none"> - Tripadvisor

9. Summary of Datasets not released as Open Data

As previously said, the project sought a balance between openness and protection, in order to comply with Open Data requirements (art. 29.3 of the GA) and, at the same time, protect those project's results whose exploitation potential could be jeopardized by an early release as Open Data (art. 29.3 and art. 27.1 of the GA), and those containing personal or sensible data.

Datasets not released as Open Data, and relative reasons for not giving access:

1. Geodatabase of irrigation channels in Sierra Nevada (Pilot #1): the database, in large part created by previous and parallels projects, contains sensitive information about the exact location and network connections of most irrigation channels in the provinces of Granada and Almeria, which may be the potential target of water contamination actions.
2. Vlach's archival and ethno-archaeological database and Geodatabase of infrastructure data in Vjosa Valley (Pilot #8): the data has been accumulated over many years from various sources and projects, presenting challenges to attribute them to a single initiative. This can potentially raise issues related to data ownership, intellectual property rights, and consent agreements with data providers if the data are used by third parties, especially for commercial purposes. Therefore, it is necessary to maintain restricted access to the geodatabase and the ethnoarchaeological archival data to ensure compliance with legal and ethical obligations regarding data ownership, as well as to mitigate possible conflicts over data attribution and intellectual property in the future.
3. Google Trends (WP3): data has not turned out to be particularly valuable and finally was not used in the analyses.
4. Historic Graves of Ireland (Pilot #9): the data collected is integrated into the whole dataset of graveyards and graves that represent the core value of the Historic Graves business model. The data is available online under a Creative Commons BY-NC-SA license but cannot be released in batch for competitiveness reasons. Each graveyard's dataset is made available, without any restriction, to the local community involved in fieldwork and data transcription, in line with the pilot's statement that "Local communities have a right to generate and own tourism-related data" (<https://doi.org/10.5281/zenodo.10979747>).