This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036458.

Data Managment Plan

D1.1

Luís Costa (PIK)
Tobias Seydewitz (PIK)

LOCALISED

Disclaimer

This report was written as part of the LOCALISED project under EC grant agreement 101036458. The information, documentation and figures available in this deliverable were written by the LOCALISED project consortium and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

How to quote this document

Costa, L. & Seydewitz, T. (2022), Data Management Plan (LOCALISED Deliverable 1.1).

General information about this Document

Project acronym	LOCALISED
Project full title	Localised decarbonisation pathways for citizens, local administrations and businesses to inform for mitigation and adaptation action
Grant Agreement no	101036458
Deliverable number	1.1
Deliverable title	Data Management Plan
Deliverable nature	Report
Dissemination level	PU
Work package and Task	WP1
Contractual delivery date	Month 6
Actual delivery date	Month 6
Authors	Luís Costa (PIK), Tobias Seydewitz (PIK)
Reviewers	Jörg Verstraete (IMP), Daniel Sánchez Muñoz (IREC)

Revision History

Version	**Date**	**Name**
Draft for review	08.02.2022	Luis Costa
Reviewed draft	28.02.2022	Luis Costa
Updated version	26.09.2023	Tobias Seydewitz

Table of Contents

List of Tables

List of Abbreviations

API	Application Programming Interface
Git	Software for tracking changes in any set of files
GDPR	General Data Protection Regulation
NUTS	Nomenclature of territorial units for statistics
IPCC	Intergovernmental Panel on Climate Change
FAIR	Findable, accessible, interoperable and reusable

Executive Summary

This deliverable sets forward the updated Data Management Plan (DMP) version for the LOCALISED project. In particular, it assesses the types and nature of data to be collected and/or generated, the quality of the generated data, and details of strategies to make data compliant with FAIR principles. Further, this updated version of the DMP introduces suggestions for appropriate licensing schemes for data and software and mitigation strategies for conflicting licensing schemas, together with planned repositories to safeguard data legacy. Further remarks on responsibilities for data publication and data protection are made. The final version of the DMP will be submitted at the later stages of the project based on this deliverable and reflecting the developments made in data production throughout the project.

1 Data Summary

This Data Management Plan presents an overview of the key data used and generated by the LOCALISED project and how it will be managed to ensure that it is FAIR – Findable, Accessible, Interoperable and Re-usable.

The information presented below will evolve and become more specific (or change) as the project progresses, and details, practicalities, and feedback from project partners and key stakeholders emerge.

The main objectives of the LOCALISED project that are most relevant to the need for a DMP are:

- To downscale energy-relevant indicators determined with the EUCalculator model to NUTS3 level and make results available to the public, as well as climate information on selected indices.
- To elaborate on two tools - the Decarbonization Profiler and the NetZero Business Consultant - that inform stakeholders on the challenges of decarbonisation.

Data will be collected and generated within each objective to serve specific purposes. For the downscaling exercise, data will be generated within the consortium using the EUCalculator model, and data will be gathered from external sources like EUROSTAT to be used as disaggregation proxies. Information on selected climate indices at the NUTS3 level will be generated from EURO-CORDEX data. In terms of software, an API will be built to facilitate the distribution of the downscaled results.

Underlying the development of the two tools in LOCALISED are particular data and software features that, due to their nature, demand the collection/generation of significant amounts of data. The most relevant for the Decarbonisation Profiler is the routine allowing users to calculate possible combinations of adaptation and mitigation options for NUTS3-regions. While in this case, much of the data will derive from the downscaling exercise previously highlighted, significant quantities of data describing relevant features of mitigation and adaptation options (e.g., costs, benefits, political feasibility, public acceptance, etc.) will be gathered from literature and existing databases of adaptation. In the case of the Business Consultant, data will be collected on the economic structure of NUTS2 regions concerning the sectoral contribution to the region's economic output. Regarding data generation, the Business Consultant will provide users with a vulnerability index generated using an adaptation of the IPCC methodology. Regarding managing code versioning and storage, the LOCALISED project will use PIK's GitLab[1].

[1] https://gitlab.pik-potsdam.de/users/sign_in

1.1 Types of data in the project

A first survey across the consortium has revealed several types and data formats to be gathered/generated, see Table 1. Data in the LOCALISED project is categorised into 3 cases:

> **Case 1:** Data is generated or collected entirely on your behalf or your institution, e.g. running your models or collecting/recording field data.
>
> **Case 2:** Data results from a re-compilation of data from third parties and potentially your data, e.g. collecting yield data from different sources like Eurostat and FAO by appending your records.
>
> **Case 3:** Data from third parties is used to compute new data, e.g. downloading climate data from CDS and generating a dataset of weather extremes.

Case 1 data will be created for country-level runs of the EUcalculator model and the gathering of businesses' decarbonisation perspectives, vulnerability, and emerging technologies. Case 2 data will be generated across WPs 2, 3, 4, 5 and 7. Specific examples of this type of data refer to data from power systems, topographic databases and infrastructure (land use, road network) or the Composition of businesses at the NUT2 level. Finally, case 3 data will be used mostly in WPs 2, 3, 4 and 5 and refers to the generation of extreme climate indices and solar and wind-power capacities per NUT-level, socio-economic data, and existing sets of adaptation/mitigation measures and existing inventories of emissions, risks, and vulnerabilities performed around the eurozone.

Regarding data formats, several but mostly orbiting around tabular (e.g., Excel, CSV) and spatially explicit information (e.g., GeoJSON, shapefiles). Data in the form of text is also expected regarding work collecting stakeholders' opinions or structuring energy policy documents or adaptation strategies.

Table 1: Envisioned data generated and collected in LOCALISED

WP	Data	Format	Comment	Open / Restrictions	Ethics considerations	Case
WP1	Impact assessment data	.xlsx, .jpeg and .doc	Obtained from project partners and stakeholders engaged in project activities	Atomic data restricted to internal use. Results at the aggregated level will be openly available.	None for data coming from partners. Data gathered from stakeholders will be based on explicit consent and a voluntary basis.	1
WP2	Country-level decarbonisation pathways	.json	Obtained from EUCalculator model	Open	None	1
WP2	Climate extremes	.csv	Processed from	Open	None	3

	of common weather variables	.geojson	EURO-CORDEX			
WP2/ WP4	European power system models	.shp .csv	Data from different open accepted models.	Open at an aggregate EU level. Local grid data can imply restrictions to be evaluated case by case during the project.	None	2
WP3	Topographic databases and infrastructure (land use, road network)	.shp	Data gathered from official APIs and web platforms	Open	None	2
WP3	Downscaled decarbonization indicators at NUTS3 level	.txt	Downscaled from EUCalculator	Open	None	2
WP3	Solar and Wind - power capacities per NUT-level	.txt	Obtained using in-house model	Open	None	3
WP3/ WP4	Socio-economic data per NUT level	.txt	Data gathered from official APIs and web platforms	Open	None	3
WP4	Collection of adaptation and mitigation measures/ options and their qualities/ KPIs/ characteristics	.xlsx .db	Data gathered from local, regional and national adaptation and mitigation plan repositories	Open after publication	None	2
WP4	Feasible mitigation and adaptation measure sets	.txt	Obtained from WP2, WP3, and WP4.1	Open	None	3
WP5	SECAP oriented SDG indicators	.xlsx	Data from various sources, SDG platform, local implementers and more	Restricted to internal use but final report open	None	2
WP5	Baseline Emission Inventory (BEI) and Risks and Vulnerabilities Assessment (RVA)	.xlsx	Calculation procedures connected to previous task and WP2 WP4	Restricted to internal use	None	3
WP5	Energy strategies	.doc	From European	Open	None	1

	linked to renovation wave and climate change		Directives (EPBD, EDD and RED), and national regulations			
WP6	Opinions on climate change and impacts on groups in 3 focus regions	.xlsx	Data generated via empirical surveys (Q Study)	Open	Survey privacy	1
WP7	Perspectives of decarbonization in Assolombarda regions	.docx	Obtained from interviews/ workshops	Restricted to internal use	None	1
WP7	Vulnerability of businesses to climate change	.doc	The exact format will be decided later	Restricted to internal use	Companies' privacy should be considered	1
WP7	Composition of businesses at the NUT2 level	.xlsx	Data from Eurostat	Open	None	2
WP7	Vulnerability index for businesses at NUTS2 level	.xlsx	Methodology developed in-house	Open	None	2
WP7	Assessment of emerging mitigation technologies and solutions	.doc	Data from expert elicitation and literature review	Open	Survey privacy	1
WPs 4,5,6, 8	feedback gathered through user consultation for co-development	.csv .txt	Obtained from interviews/ workshops/ questionnaires	Restricted to internal use	None	1
WP9	Scientific materials from speakers and experts	.doc	Participation in webinars/ seminars	Open	Previous consent will be required to the participants	1

The data generated/gathered in the LOCALISED project is expected to be most useful to municipalities, businesses and citizens concerned with or promoting decarbonization and adaptation activities. The data will also benefit cities participating in the project and associated partners. In addition, data is also envisioned to be useful for the scientific community dealing with energy-relevant topics such as energy modelling or regional development. The data generated by the project in the tasks outlined above (along with those from other activities and tasks) will be made available principally through three publicly accessible outlets: a database of downscaled decarbonisation indicators at NUTS3-level, the Decarbonization Profiler and the NetZero Business Consultant. To facilitate the transfer and integration of the data generated between specific project

outputs and the LOCALISED tools via their APIs (see Table 2), the common data format JSON will be evaluated for the necessary flexibility.

2 FAIR data

2.1 Making data findable, including provisions for metadata

Specific keyword and metadata terms and conventions will be defined as the data is produced and online tools are created. Naming conventions shall be developed on project levels and will follow standards (e.g. ISO) where such standards are applicable.

The naming convention will be documented in the metadata for the dataset. For generic repositories, the search keywords will be extracted from the metadata and provided as part of the descriptive metadata for the dataset. If there is a use of subject-specific repositories, the search keywords will follow the standard utilised by the given repository, e.g. following a specific ontology. The project will adhere to the best practices of search engine optimisation (SEO) best practices for each LOCALISED tool. This means they will be easily discoverable by search engines, using HTML meta tags to describe the content of the tools in a clear and easily understandable manner. The project will heavily advertise the presence and location of the online tools.

Each data owner is responsible for deciding the file naming conventions and including a clear version-control guideline. Version control will follow the file naming conventions that include a version control in the file name.

2.1 Making data openly accessible

2.1.1 Accessibility for project partners

Unless ethical issues arise, all data generated in the LOCALISED project will be available for the project partners throughout their lifetime. Applications developed within the LOCALISED project are shared via repositories published on PIK's GitLab instance. In case of access restrictions, these repositories are mirrored on GitHub. Small datasets required to run the applications are embedded in the repository, and larger datasets are shared through the PIK cloud service.

2.1.2 Accessibility for the public

The results of the LOCALISED project will be fully accessible to the public unless ethical issues regarding particular data arise or constraints in the background section of the consortia agreement apply. Data provided via the LOCALISED planned APIs and tools will be fully accessible to the public. The datasets contained within and presented by each tool will be viewable directly within the tools themselves or extractable in a widely usable format (e.g., xlsx) for further analysis. To preserve transparency, the code for each of the online tools, as security concerns allow, see Table 2 WP4 and WP8, will be made open-source and accessible via public repositories in a running instance of GitLab

on the servers of the PIK. Further, the repositories are mirrored to public repositories on GitHub if required.

Table 2: Methods and software to access the data and generated results of LOCALISED

Work Package	Methods	Software	Open source
WP2	Percentile-based extremes	Python	Yes
WP3	Data disaggregation tool	Java/Python/SQL	No
WP3	API to access the energy/emission database		Yes
WP4	Python-based, algorithmic ranking of feasible measure sets	API interfacing data to the lightweight model	Yes
WP4	Selection, combination and curation of model output	API interfacing results from the lightweight model to the tools	Yes
WP4	Database development	.db; .xlsx	Yes
WP4	GIS Vulnerability mapping (Summation; PCA)	ArcGIS; QGIS	Yes
WP5	Calculation of SDG indicators to implement it on the platform.	.xlsx	Yes
WP6	Q Study data collection and analysis	Q Method Software	No
WP8	Frontend accessing project results		Yes

The data made open by the LOCALISED project will be stored under established repositories to preserve its long-term usability – see Table 3. The repositories comprise those oriented to the scientific community but also a dedicated LOCALISED data-sharing platform and webpage. Further, in LOCALISED, various repositories are used to meet the requirements for publishing the project results, which a one-source repository can not fulfil. Datasets are published in the Zenodo LOCALISED community, which assigns a unique identifier (DOI) to each release, making it citable and long-term available. Additionally, it is possible to publish large datasets, which is not feasible via a Git-frontend service. The code developed in the project is released on the previously mentioned Git-frontend services because Git provides the best service in terms of code sharing, version control and collaborative work. Git publishing enables Stakeholders interested in the project code to clone the developed code easily or to review it. Subject to evaluation and confirmation, some of the outputs of LOCALISED are of potential interest to the Covenant of Mayors (CoM), given its mission in facilitating cities with elaborating their decarbonisation plans. Accordingly, LOCALISED will meet 2024

representatives of the CoM platform and evaluate the possibility of publishing relevant project data and outcomes via the CoM platforms.

Table 3: Overview of data repositories for the project

Data repository	Comment	For which data?
Zenodo[2]	General-purpose open-access repository developed under the European OpenAIRE program and operated by CERN.	All data released for public use
PIK LOCALISED cloud folder	Restricted to project participants	Internal project data and development
Localised Data-Sharing Platform	Implemented on the FZJ OpenStack Cloud. General Access via REST API.	Downscaled energy indicators to NUTS 3 level for all of Europe
Git https://git-scm.com/	Free and open source distributed version control system	For project APIs and code
LOCALISED webpage[3]	Accessible to all although particular restriction on some data might apply	Communication-related activities and stakeholder engagement

2.2 Making data interoperable

For LOCALISED, proprietary data formats should be, as much as possible, replaced by their open counterparts, e.g., replacing ESRI-Shapefile with GeoJSON or replacing Microsoft Excel files as OpenOffice files. To ensure interoperability of data files between different operating systems, save all data with UTF-8 encoding and OSX/Linux line endings using only line feed (LF, "\n").

LOCALISED prefers, if feasible, to use SQLite if data is stored in a database management system (DBMS) due to the ease of exchange of database files between project partners or shipping to end-users. Data stored in a DBMS must be normalised and preferably distributed with an entity relationship diagram. Further, include database documentation, explaining the stored attributes per table in short phrases.

Applications developed during the project require a readme explaining the dependencies and requirements to run the software and installation guidelines. Document application code according to the documentation style suggestions of the programming language. Additionally, format source code files according to the syntax style guidelines of the language. In LOCALISED, it is preferred to prove the validity of the code by providing unit tests. Bug tracking and user feedback are organised through GitLab or GitHub's issues trackers.

[2] https://zenodo.org/communities/localised/
[3] https://www.localised-project.eu/

2.3 Increase data re-use

2.3.1 Licensing

For data that is planned to undergo public release licensing, the LOCALISED project will adopt the Creative Commons Attribution International Public Licence CC BY 4.0 or later, and the Open Data Commons Attribution Licence ODC-By v1.0 or later. Both licences do not have a copyleft clause and are compatible and comparable. They allow you to share (copy, redistribute and use the data), to create (produce works from the data) and to adapt (modify, transform and build upon the data) as long as you give appropriate credit.

For code planned to undergo public release, the GNU Affero GPL licence (version 3 or later) will be adopted. It is fully compatible with the commonly used standard GNU GPL licence (version 3 or later), but in addition, it covers the use case of "Software-as-a-Service" (SaaS) by also requiring, in these cases, the publication of the corresponding source code. This becomes relevant if someone provides access to a model as an online service by running it on a web server. Note that the suggested licence is "copyleft". This means that a derivative of a product has to be published under the same conditions. This basically makes sure that code published under such an Open Source licence will also remain open in the future. This gives free software developers an advantage over proprietary developers: software that they can use, while proprietary developers cannot use it without releasing the source as well.

In LOCALISED, partners could see the urge to sublicense code or data provided by third parties. Here, it applies the principle that the sublicense can only be as open or closed as the top license allows. If a partner needs to sublicense, priority is given to licenses that comply with the set of licenses described in this data management plan. The minimum license requirement is that the product can be released as open source.

2.3.2 Data quality

The data provided via the LOCALISED tools will be made available to the public once enough confidence of its robustness is demonstrated by the data producer (partner in charge) and generally accepted by the team, or in case the data that has been part of scientific work fit for publication. This ensures that the data is subject to additional quality control from the review process. After data reuse is allowed, third parties may freely use it for non-commercial and educational purposes.

Further, each data producer in LOCALISED must record the quality of the produced data based on the following criteria - see Table 4:

> **High:** The produced data is sourced from a peer-reviewed publication or an external source, providing a transparent quality report. Further, this quality level comprises all data cases listed in Section 1.1 that will be published in peer-

reviewed journals or reviewed by a LOCALISED partner, or a data quality report is publically available.
Moderate: The produced data is sourced from a peer-reviewed publication or an external source, and only minimal processing is involved in producing the output, e.g. aggregations or descriptive statistics.
Low: Applies to all heavily processed data without any review process involved.

Table 4: Overview of produced data quality

Work Package	Data	Quality
WP2	Hazards and Impacts	Medium
WP4	Measures' implementation data	Moderate for mitigation measures Moderate - Low for adaptation measures
WP4	Measures' feasibility data	Moderate
WP4	Measures' descriptive data	High
WP6	Descriptive demographic data on vulnerable groups	High
WP	Opinions and subjective self-assessment, sourced via surveys	Low
WP6	Energy poverty indicators data	High - moderate
WP7	Emerging decarbonisation solutions for businesses	High
WP7	EU business vulnerability index	Medium

3 Allocation of resources

The costs for making the data generated by the project 'FAIR' will be minor and covered as part of the budget assigned to the project partners responsible for producing the data and associated tools. The responsibilities for data management rest in the first instance with the project partners responsible for generating or compiling it and PIK as project coordinator. Preserving data in the long-term, so after the end of the LOCALISED project, will be handled by sharing code and data through the corresponding repository providers - see Table 3. This can be done without any considerable additional resources. The data handling for the tools is related to the sustainability of the tools and is under discussion in the consortium at the time of writing this version of the DMP. Currently, it is planned that the Business Consultant will be available through CMCC services. Therefore, the data management within the Business Consultant would stay in the responsibility of CMCC and will be covered by their resources. For the Decarbonisation Profiler, options are discussed, from licensing to transferring the tool to institutional actors; however, no specific decisions were taken and probably can only be planned towards the end of the LOCALISED project once users' interest has been tested.

Therefore, the necessary resources for data management currently need to be estimated. This can only be done in the final version of this deliverable.

4 Ethical aspects

In LOCALISED PIK serves as a legal and ethics partner. As such, PIK is dedicated to ensuring that legal requirements are complied with throughout the project, particularly regarding data protection law and that data processing activities are ethical. Data protection law may, in certain cases, prohibit or limit the possibility of making data openly accessible; see data classified as Case 1 in section 1.1. In these cases, ethical considerations will be treated following the outlines in D1.2 "Ethical Requirements" and supervised by the "Intellectual Property Rights Committee" of LOCALISED.

Based on a sound analysis of the ethical and legal frameworks applicable to the project, PIK will elaborate guidelines (such as the LOCALISED Privacy Policy) and set up appropriate legal and compliance strategies to ensure that data subjects and data owners remain in control of their personal data and subsequent use. That data is processed within the LOCALISED project in compliance with subjects' legal rights, particularly those bestowed by GDPR. This will apply to all data collected and processed within the project, including the accelerator/residential program, where data are gathered via questionnaires and interviews, the data related to the open calls, and data from dissemination activities, including personal data such as contact details from participants and stakeholders.

As such, the LOCALISED project will show deep respect for privacy and data protection as both a legal requirement and an ethical standard. All project partners regard personal data protection obligations as an ethical best practice standard in this context.

www.localised-project.eu