LAMASUS
Land Management for Sustainability

D10.2 Data Management Plan

Funded by the European Union

Date
27.02.2023

Grant Agreement	101060423
Call identifier	HORIZON-CL6-2021-GOVERNANCE-01
Project full title	Land Use and Management Modelling for Sustainable Governance
Work package	WP10
Due date	28/02/2023
Deliverable lead	Tamás Krisztin,
Authors	Tamás Krisztin, Albert Brouwer, Ellen Tan

Abstract
The Data Management Plan (DMP) serves to ensure that data used in and produced by the LAMASUS project are findable, accessible, interoperable, and reusable (FAIR). This data management plan gives an overview on the approach LAMASUS takes to ensure the FAIR principles. The publication of research data is closely linked to the publication of journal articles or other outlets for such results. In general, LAMASUS published open access articles and applies the same policy for the underlying data.

Keywords
Data Management Plan (DMP), FAIR principles, findable, accessible, interoperable, reusable.

INTERNAL TECHNICAL AUDITOR	NAME (BENEFICIARY SHORT NAME)	DATE OF APPROVAL
Task leader	Albert Brouwer (IIASA)	28/02/2023
WP leader	Tamás Krisztin (IIASA)	28/02/2023
Coordinator	Petr Havlík (IIASA)	28/02/2023
Project Office	Elija Wiedemann (ARTTIC)	28/02/2023

Dissemination level

PU	Public, will be published on CORDIS	✓
SEN	Sensitive, only for members of the Consortium (including the EC services)	

Nature of the deliverable *	R

Table of Contents

List of Tables

Abbreviations

DG AGRI	Directorate General for Agriculture
DMP	Data Management Plan
EC	European Commission
EU	European Union
FADN	Farm Accountancy Data Network database
FAIR	Findable, accessible, interoperable, and reusable
GDPR	General Data Protection Regulation
LUM	Land use management
NUTS	Nomenclature of Territorial Units for Statistics
OA	Open Access
WP	Work Package

1. Introduction

This Data Management Plan (DMP) describes how the LAMASUS project aims to ensure that all project data and its processes conform to Horizon Europe's open access policy and recommendations on open research data, as well as the EU General Data Protection Regulation (GDPR). It outlines the strategies to ensure the data used and produced by the project are findable, accessible, interoperable, and reusable (FAIR) both during and after the project life.

The DMP addresses all aspects of data management, with particular attention to:

- What data will be collected, generated and/or processed by the project.
- What methodology and standards will be applied.
- Whether and how the data will be shared/open and FAIR.
- How the data will be handled, retained, and preserved, during and after the project.

This DMP covers all stages of the information cycle, including the management of external inputs to the project (e.g., Work Package (WP)2 and WP3), the databases produced (WP2 and WP3), the intermediate products exchanged internally (all WPs), and the data results of simulations and scenarios produced (e.g., WP5, WP6, WP7, WP8). It includes guidelines and requirements for the documentation and preservation of all the information used and produced, as well as the treatment of personal data of both internal and external stakeholders (e.g., WP1, WP10), and data security and GDPR considerations.

The DMP is considered a living document and is subject to updates whenever significant changes arise and over time with periodic project evaluations.

2. Data Summary

All data used within the LAMASUS project will contribute to the project objectives. One key objective of the LAMASUS project is to create a harmonized, high-resolution mapping and monitoring of land use management (LUM) and EU agricultural and forest related policies. For this effort we combine remote-sensing and European statistical data from 1990 to 2020 and update it moving forward, which is important for policy modelling purposes. In addition, we include spatially explicit LUM drivers of European agricultural and forestry policies and agricultural and forestry related direct payments and quantitative proxies of EU-wide policies at the Nomenclature of Territorial Units for Statistics (NUTS) 2/3 levels, and national levels from 2000 to 2020.

Data and policy models will be made publicly accessible through the LAMASUS Portal. The Portal will also provide analytic and visualization capabilities for analytic and policy making purposes and incorporate on the ground feedback to assess the accuracy of results.

2.1. DATA

The following categories of data will be collected, utilized, and processed during the project:

- Demographic information: about the project partners, other interested parties, and users of our tools and outputs (e.g., LAMASUS website).
- Land use and land cover data: diverse sources of historical data as well as future projections and in-situ related data on different aspects of knowledge on land use practices, management, and data derived from remote sensing sources and official statistical databases.
- Other spatially explicit data: data on soils, habitats, protected areas, agricultural statistics, crop parcels and pesticide use.
- Climate data: different sources of climate information, including historical weather observations, processed climate data (e.g., interpolated, reanalyzed), based on both in-situ and remote observations.
- EU-wide Farm Accountancy Data Network database (FADN) data: annual data on farm income and business activities. The FADN survey catalogue are extracted from inventory, cash book or journal kept by the farmer or field officer and are grouped into 13 tables. FADN consists of bookkeeping data for a subset of farms.
- Integrated data in the LAMASUS Portal: different sources and types of socioeconomic and environmental data available for the EU. Such information includes spatially explicit projections of land management that combine several of the data sources listed above.
- Data generated by participatory activities and stakeholder engagement: Any data produced through interviews, case studies and online surveys will be generated and contextualized, and only made publicly available in an anonymized format and under informed consent from all involved parties.

Data types to be managed in the project

As part of the project, the consortium will manage different types of data with varying degrees of sensitivity to disclosure and terms of use. The four general groups of data that will be managed can be grouped into the following broad categories:

- Publicly and non-fully publicly available data sources (e.g., national agencies, international initiatives, and research institutes).
- Peer-reviewed and non-peer reviewed literature (e.g., scientific articles, technical reports)
- User input data (e.g., interviews, questionnaires)
- Personal data: data that relates to an identified or identifiable living individual, or different pieces of information, which combined can lead to the identification of a particular person.
- Model-based data (e.g., modeled species distributions).

Volume of the data

To assess the characteristics and volume of the data to be collected, processed, and managed during the project, a survey was conducted to help provide a more accurate estimate on data storage and accessibility needs. Because the data collection is only starting for some work packages, the exact volume of data will be amended during subsequent versions of the DMP.

The results of the survey are summarized per work package and beneficiary in Table 1 (re-used data) and Table 2 (generated data). The first column provides information on the work package, the second on the data collecting beneficiary. The third column describes the type and format of the data to be collected, while the fourth column indicates the beneficiary's estimate of the expected data size. The final column containing a description of the dataset's origin or provenance if applicable.

Table 1: Overview of the data reuse activities and expected size of the data by work package and main beneficiary.

WP	BENEFICIARY	TYPE AND FORMAT OF THE DATA	EXPECTED SIZE	ORIGIN / PROVENANCE
1	WIFO, WUR	Policy documents, research papers	Several kilobyte	
2	IIASA	Raster data, vector data, point data and geotagged photos, including: • Remotely sensed data products on land use and land cover • Other spatially explicit information, e.g., soils, habitats, agricultural statistics, crop parcels, etc. • In situ data on land use, land cover, cropland, e.g., LUCAS	Several terabyte	Copernicus, national mapping agencies, European and national environmental agencies, European and national space agencies, European Data Cube, Global Forest Watch, Zenodo, Eurostat, national statistical offices, national forest inventories, DG-AGRI
3	TI	FADN payments data	Several hundred megabyte	DG AGRI, FADN database
4	IIASA, INRAE, UW, VUA, UB, WIFO, RUR, WUR, UV, ZHAW	EU statistical FADN, climate, soil and land use data French level pesticide, land parcel and climate data: Norway farm statistical data EU spatial drivers of land-use and management (e.g. slope, soil types, etc) Socio-economic data at the NUTS2/3 level	Several gigabyte	DG AGRI, FADN database, JRC, Eurostat, Corine, LUCAS, Land Parcel Identification System (LPIS)
5	INRAE, UPS, IIASA, PBL	FLUXNET and Integrated Carbon Observation System data. Inventory data for	Several terabyte	Existing literature, data obtained from research institutions, global initiatives, and national authorities (e.g., European

WP	BENEFICIARY	TYPE AND FORMAT OF THE DATA	EXPECTED SIZE	ORIGIN / PROVENANCE
		- soil C stocks are available from World Soil Information Service (WoSIS) database - soil C stock changes from different meta-analysis published in scientific journals and from the Crop Research Institute in Prague, Czech Republic - GPP/NPP/biomass also from different inventories - MODIS data will come from the corresponding web-archives.		Environmental Agency, EEA). Information will be compiled from publicly and non-fully publicly available datasets, such as the Global Biodiversity Information Facility (GBIF), iNaturalist, European Vegetation Archive (EVA) and eBird, for all dimensions and scales of biodiversity, including species and their traits, habitats, and ecosystem services.
6	VUA, RUR, ZHAW, BOKU	FADN data including farm assets, crops, debts, farm classification, general information, inputs, labor, livestock, reference, and subsidies. Data on resources endowments and observed land use can be obtained from these data groups for the modelling approach.	Several gigabyte	FADN database
7	IIASA, Eurocare, WUR, PBL	Spatial explicit layers of FADN variables and NUTS2 aggregations.GLOBIOM, CAPRI, MAGNET, IMAGE model data (land use and management data, projections of socio-economic variables, prices)	Several gigabyte	FADN database GLOBIOM, CAPRI, MAGNET, IMAGE data repositories

Table 2: Overview of the data generation activities and expected size of the data by work package and main beneficiary

WP	BENEFICIARY	TYPE AND FORMAT OF THE DATA	EXPECTED SIZE	ORIGIN / PROVENANCE
1	WIFO, WU	Meeting notes and interviews with stakeholders, contact data of external / internal stakeholders.	Several megabyte	WP1
2	IIASA	Annual raster geodatabase (1990 to 2024) of land use and management and vector aggregations to NUTS level; raster of potential areas for peatland rewetting.	Few gigabyte	WP2; this will be a derived or added value product, so based on many existing input data sets including FADN data
3	TI	Spatial explicit layers of FADN variables and NUTS2 aggregation used to generate a synthetic farm population. Regional farm structure are determined based on farm types, location information, land and animal use, yields, subsidies, and total utilized agricultural area.	Few gigabyte	WP3; FADN data base
4	INRAE	Estimated elasticities of pesticide use with respect to weather variables change. Estimated elasticities of land use with respect to rents and climate variables.	Few hundred megabyte	WP4
5	INRAE, UPS	Knowledge base of climate impacts of different LUM classes and transitions. Maps of potential climate impact of LUM classes and LUM transitions	Few gigabyte	Produced from model simulation within Task 5.1, re-use of certain functions from G4M

WP	BENEFICIARY	TYPE AND FORMAT OF THE DATA	EXPECTED SIZE	ORIGIN / PROVENANCE
				From application to response functions from knowledge base to LUM maps produced in WP2
6	VUA	FADN data are used for various purposes: • improve the models intensity data • derive probabilistic inferences between farm and farmer characteristics and their adoption of practices and land management techniques • create an organic farming map of Europe • create a projected land systems raster map of EU27 + Norway and Switzerland for 2030/2050	Few gigabyte	WP6
7	IIASA	Spatial explicit layers of FADN variables and NUTS2 aggregations.	Several hundred megabyte	WP5
8	PBL, IIASA, VUA, WUR	Scenario projections from the main ex-ante models from WP6 and WP7.	Several hundred gigabyte	WP7; produced by ex-ante modelling teams
9	AI	Text and images from project work plan and from (scientific as well as general interest) publications arising from the project.	Up to 1GB total	Consortium members, Industrial Advisory board, public databases
10	IIASA	Contact databases of the project	Up to 1MB	WP10

2.2. ACQUISITION OF REUSE PERMISSION

LAMASUS project partners should personally contact data owners as a method to secure reuse permission, unless the reuse and acquisition of data is explicitly permitted (such as with official statistical data sources from e.g. Eurostat). Permissions must be gathered, documented, and stored by partners along with the relevant data when the data is the property of an external research institution, business organization or another type of legal entity. The following information should be documented and stored on the LAMASUS SharePoint in the relevant WP folder:

i) Information about the data:
   - Title
   - Author(s)
   - Digital Object Identifier (DOI, if available)
   - Short description (~100 words)
   - License
   - Website where data resides (if appropriate)
   - Contact information of the organisation owning the data (including name and physical address of organization)
   - Contact information of the person giving access to the data (including name of contact person, e-mail, and phone number)

ii) Information about the data requester:
   - Contact information of person requesting permission (including name, e-mail, phone number, physical address, and organization)
   - Dates when permission to use data was requested and granted
   - Specific purpose(s), i.e., how the data will be used, for which task(s) and WP(s)

iii) Conditions for which the permission to use the data was granted

Copy of the correspondence or letter where the permission was granted. These four elements listed above should be combined into a single PDF document and uploaded to the LAMASUS SharePoint using the following naming convention: Permission_YY-MM-DD_Title-of-the-data.pdf, where the date represents the day when the permission was granted.

Information about data reuse may also be featured in respective scientific publications that present results obtained through reused data in accordance with the policies of the scientific publisher (e.g., journal, conference proceedings, book).

3. FAIR data

The so-called FAIR guidelines (for a reference see "FAIR Guiding Principles for the Management and Stewardship of Scientific Data", Wilkinson et al., 2016) were intended to provide guidelines for improving the findability, accessibility, interoperability, and reuse of digital assets. These stylized principles refer to three main types of entities: data (or any digital object), metadata (information about that digital object) and the infrastructure that will support and maintain these first two entities over time. When considering FAIR principles, one should share the data and its corresponding documentation in such a way

that it reaches as many people as possible. The most efficient mechanism to achieve this is to make the data and their descriptions available online in a persistent manner.

It is possible that due to privacy issues (e.g., farm-level, or personal data), sensitive information (e.g., estimated location of farms), or even the size of data files, it may be difficult in practice to share some of the data used and produced by LAMASUS. But, at a minimum, each data source should document what data, when, and how they were used or produced in the project. These answers will be made available on a publicly accessible website. This ought to help stakeholders and the wider public to understand what was done and how they could potentially build on it. Within the LAMASUS project it is foreseen that data will be shared using the LAMASUS data explorer, hosted on the project website.

3.1. Making data findable, including provisions for metadata

Research projects such as LAMASUS use and generate a large amount of data during their activities, and metadata should be used to describe this data to facilitate its management and make it FAIR. Metadata provides descriptive, structural, and source information about the data, such as its content, organizational relationships, management rights, and intellectual property rights.

Descriptive metadata serves to discover and identify the content of the data, while structural metadata describes the organizational relationships of the data. Source metadata includes information about management rights, intellectual property, and intellectual property rights (Riley, 2017). Proper documentation of metadata is essential to ensure that the data is properly managed, as it can provide a basis for understanding the context and purpose of the data and how it can be used. Additionally, metadata facilitates the search and analysis of the data, as well as the standardization and change management of the data, resulting in improved information management and evaluation processes.

Processed and generated data will be made accessible to the public in compliance with the obligations for data protection by the GDPR and restrictions imposed by the data providers. Zenodo uploads will be used to provide a Digital Object Identifier (DOI) to releases of public LAMASUS datasets. The data sets will be tagged as part of the LAMASUS Zenodo community.

The generation of appropriate and sufficient metadata will support the FAIR nature of data. Each dataset will be findable and searchable through the added Zenodo-format meta data. The Zenodo search functionality enhances the probability of identification using keywords and community grouping. Both individual release versions and an overall series of releases of a dataset can be provided with a DOI by Zenodo.

Additional metadata files may be included with the data set as appropriate for the particular kind of data or research field. As these files will be included in a repository (see below) and are downloadable from Zenodo, they too can be in principle indexed and harvested, given knowledge of their presence and naming. Both descriptive, structural, and source metadata will be used within the different research stages of the LAMASUS project.

To comply with the principles of open data, partners will be encouraged to structure all gridded multidimensional spatial data produced by the project in netCDF format and ideally document it using established metadata convention standards, such as the Climate and

Forecast (CF) Metadata Convention for netCDF. Nonspatial data should be documented providing the DataCite's Metadata Schema minimum terms.

3.2. MAKING DATA ACCESSIBLE

To ensure open access data, LAMASUS will digitally archive and compile the datasets generated by the project and their associated metadata in appropriate data repositories. For this we request all partners and collaborators to deposit the data resulting from all project activities in either Zenodo or Git reference repositories. The Zenodo REST API allows automated harvesting of data and metadata. Automated downloads are possible via the https protocol.

The evolution of datasets and the associated metadata will be tightly managed via Git. Guidance documentation on how to do so will be provided to LAMASUS partners. Git repositories for publishing and sharing by default will be hosted under the LAMASUS group on the IIASA GitLab server. This server supports the Git Large File Storage (LFS) extension for storing large data files by reference in a storage backend. The repositories are referenced from the metadata of associated Zenodo uploads that correspond to dataset releases. The repositories can also be connected to Git repositories hosted elsewhere (Git remotes) for processing or editing. Specific commits of repositories can be included by reference (Git submodules) in other Git repositories.

Owners of each dataset can choose a custom workflow for evolving the data and selectively sharing it, making it public, or keeping it private as required for their specific use case. This is done by reconfiguring the access settings of the hosted Git repositories. Partners will be provided with guidance on how to set up a customized Git workflow. While data is under testing and being evolved, the data may not be public at the discretion of the data owner. On release of a well-tested and completed dataset, the data will be made public. Releases of data and associated code are made through the GitLab release procedure or GitHub release management.

For data generated as part of the project deliverables, no embargo will be applied to give time to publish or to seek protection of intellectual property. There is no need for a data access committee.

For private repositories, particular users can be selectively given access based via their GitLab/GitHub accounts. Access rights can be configured per user. Users with access can browse a repository with a web browser and clone the repository with a Git client using one of the protocols that Git supports (https, ssh, ...). The identity of users is verified by verifying that they have control over an institutional email address tied to their person.

Within the consortium it is recommended that partners use data formats that are lightweight, non-proprietary, and as openly accessible as possible. Examples include the use of comma-separated value files (CSV) for tabular data and the use of netCDF (Network Common Data Form), GML (OpenGIS® Geography Markup Language 2007), GeoJSON (GeoJSON Format Specification 2015), GeoPackages (GPKG) and other similar standards for exchanging geospatial data. All spatial data of the project should follow the EU directive INSPIRE (Infrastructure for Spatial Information in Europe).

Meta data will be included in the Git repository containing or referencing the data. The data on the LAMASUS GitLab instance hosted by IIASA will remain available as long as it is

contractually required by the EC. Zenodo is hosted by CERN and the uploads thereon also remain available within the bounds of CERN institutional longevity.

DOIs or other persistent identifiers (e.g., “stable URLs”) to the data deposited in repositories, as well as the name of the repository, should always be included in the paper using or describing that data resource.

LAMASUS Portal

All major datasets generated during the LAMASUS project will be hosted on the LAMASUS portal. The LAMASUS Portal includes the following operational components:

1. The **Database** hosts public and confidential (to be accessed by project participants only) databases using an API for automatised access.
2. The **Model Repository** hosts model software codes and documentation with example and instructions on how to update the scenarios and run the projections.
3. The **High-resolution Scenario Validator** visualizes interim modelling results from the Database and provides an interface for accuracy feedback.
4. The **Data and Scenario Explorer** provides a web-based interface that allows for in-depth visualizations of projected national, and supra-national times series data and long-term scenario outputs, and, allow for selection and download of the Data and Model Repository content.
5. The **Land Policy Dashboard** visualizes data and model projections on policy relevant maps, from NUTS2 to country level resolution, to allow for comparisons across spatial and/or variable dimensions.

Components 3. to 5. are one seamless web-based product, where the user can switch between the Scenario Validator, Data and Scenario Explorer, and Land Policy Dashboard.

3.3. MAKING DATA INTEROPERABLE

The LAMASUS guidance is to opt for well-supported and widely-used open data formats and metadata standards where possible. Metadata standards will be chosen by exploring the Digital Curation Centre Metadata Standards and the Metadata Standards catalog. Specifically, the interoperability of the LAMASUS project and the data managed and created in it will be achieved through the following aspects:

1. The communicative interoperability of the organizations participating in the project, achieved through the understanding of the data requirements and consensus among different institutions of the consortium to be able to manage the data in a standardized way.
2. The semantic interoperability, by clearly defining the structures of the data to be generated by LAMASUS and how they will be codified. This includes using established community data interchange guidelines (such as the Agricultural Model Intercomparison standards).
3. For the collection, documentation, and storage of the different gridded spatial-temporal scenarios the project will rely on the netCDF data structure, which is particularly effective in structuring, storing, and distributing gridded spatial data.
4. The technical interoperability will be achieved through the implementation of the solutions and platforms required in the different phases of the project, such as the

LAMASUS Web portal, which is expected to allow users to visualize and summarize the project data outputs.

3.4. INCREASE DATA RE-USE

Access to data deliverables in the LAMASUS project will be achieved by making them available online, so that they can be downloaded, analyzed, reused, and cited by people and organizations other than the creators of the data. To achieve access that complies with the guidelines of the FAIR principles, multiple aspects shall be considered: data hosting, metadata documentation, publication licenses for datasets, and software.

Data hosting

Final data created and modified by LAMASUS will be made available through project specific repositories that provide DOIs as a mechanism. Through this principle data should be available for download for at least the next 10 years, depending on the commitments of the hosting institution.

Metadata documentation

The consortium seeks to achieve a balance between necessity to document data to specific standards and the complexity linked to doing so (and associated burden on consortium partners).

Nonspatial data should be at a minimum documented providing the DataCite's Metadata Schema minimum terms:

- Resource type (publication, poster, presentation, dataset, image, software, workflow, other)
- DOI
- Publication date (YYYY-MM-DD)
- Title
- Creator(s) (i.e., family name, given name)
- Creator(s) affiliation (institution)
- Description
- Access right (i.e., open, embargoed, restricted, closed)
- License type (Creative Commons license)

All gridded multidimensional geospatial data produced by the project should be structured in netCDF format or comparable format and documented using the Climate and Forecast (CF) Metadata Conventions.

Publication licenses for datasets

LAMASUS is fully committed to the Open Data (OD) quota and Open Access and the consortium will adopt the Horizon Europe Open Science Principles. All project data outputs and deliverables should be published under Open Data Commons Attribution License CC-BY 4.0 license at a minimum, with even less – but not more - restrictive licenses (e.g., CC-0) being upon the discretion of data creators. Each partner has carefully budgeted to ensure publications can be published in open access journals. Project-related reports and policy briefs will be made publicly available through the LAMASUS website (www.LAMASUS.eu)

Software
Each hosted repository has a unique URL that makes it downloadable via a Git client and browsable via a web browser. A permissive license such as CC BY will be the default, with licensing specializations up to the discretion of data owners. To the maximum extent possible, data sets will be made publicly accessible, but will only be released after undergoing refinement and testing to optimize the data quality. Release management will be done as part of the management workflow, with the public hosted repository typically receiving only well-tested updates, and experimental work confined to local repositories or privately hosted repositories.

The custom nature of some datasets makes re-use only realistic when the public or partners are provided with the code that processes the data, or code examples that process the data. In such cases the guidance is to release such code together with the data. This can be done in the same Git repository or, alternatively, via an umbrella repository that contains the code that operates on the dataset, and references a specific version of the dataset — by default hosted on the LAMSUS GitLab instance — as a Git submodule.

When such an umbrella repository is hosted on GitHub, an alternative release workflow with automatic upload and DOI provision can be chosen by making use of the Zenodo GitHub integration. Zenodo uploads can also be automated via GitLab CI/CD.

Once released, datasets will remain available indefinitely. In case of discontinuities in hosting, the distributed nature of Git gives good assurance that the data can easily be re-hosted elsewhere.

4. Other research outputs

LAMASUS has a contractual obligation to make available all the project outputs and deliverables of the projects under open access with certain exceptions defined in the Grant Agreement. The list of other output categories presented below do not constitute an exhaustive list of all the existing possibilities. We have tried to highlight the most important categories within LAMASUS.

Peer review research articles and pre-prints
This category includes peer-reviewed articles that report the original results of a research project and are published in an academic journal; review and case study articles that summarize the state of the art regarding research on a given topic and can provide an overview or report on active debates in a field, gaps in knowledge, or predict the direction a field will take in the future; and case studies articles that provide in-depth investigations that analyze and explore specific aspects of a system to explain the underlying concepts, causal links and implications that a case topic has in its real context evaluating the response of a system under real conditions. In addition to peer-reviewed articles are preprints, which are complete drafts of a research manuscript that are uploaded and shared in a public repository (ArXiv, Zenodo) prior to formal peer review.

Peer-reviewed scientific articles must follow the Gold OA model, i.e. they must be immediately available in OA mode under at least the LAMASUS default license (see above, CC-BY 4.0), with associated costs included in each partner's budget. Non-peer-reviewed

documents must also be open access, and these include reports, preprints, policymaker briefs, etc.

Project related reports, deliverables, and policy briefs
Reports are generally non-peer reviewed and can take many forms and serve different functions within a project (e.g., a short executive summary report that aims to inform policy, or internal project reports or strategies). They are usually written by one or more individuals, and often receive feedback from a wider project consortium group. These will be publicly available as far as possible and hosted for the public on the LAMASUS website.

Software code
Model code and data processing code are two obvious additional research outputs. Part of the intent of choosing a Git-based data management approach as outlined above is that code can be managed with the very same tools and techniques, and even as part of the same workflow. This reduces fragmentation and learning and makes it easier to keep code and data in sync.

5. Allocation of resources

Direct and indirect costs associated with the dissemination of LAMASUS results following FAIR principles have been considered in the project Grant Agreement. The relevant partners have attributed budget to making research results openly available. Costs for Open Access publications are budgeted for by the contributing partners for the duration of the project. In some cases, the libraries of partner universities have agreements with journals to cover Open Access fees, and where necessary partners will post versions of published manuscripts in self-archived online repositories. The costs associated with the long-term maintenance of data archives will be covered by IIASA as project-in-kind contributions, including the maintenance of the server space where all the data produced by the project will reside after the project ends.

Zenodo, the LAMASUS GitLab instance at IIASA, and GitHub are all backed by redundant storage and backed up. The use of alternative platforms such as Zenodo is without costs and uploaded documents and data are guaranteed to be made available for at least 10 years. Datasets associated with published peer-reviewed articles will remain available on the homepages of the respective open-access journals.

The responsibility of data management lies with the LAMASUS beneficiaries, specifically the owners of the respective datasets.

6. Data Security and protection of personal data

The responsibility to ensure data (including personal data) collection, storage, protection, retention, and destruction according to relevant regulations and Article 15.2 of the LAMASUS Grant Agreement ("Data processing by the beneficiaries") lies with the institutions or departments of each of the partners.

For truly sensitive datasets, a network filesystem with select institution-local user access is available at the institutions of participating partners. For these instances, a Git-based workflow will have to be mostly abandoned and laborious manual management and coordinate will have to take its place with no option for release and publishing. However, in the context of LAMASUS it is not envisioned that personal/sensitive data with restricted access will have to be distributed.

For non-sensitive data hosted Git repositories will be used as described above. Git repositories can easily be cloned for processing elsewhere or on fast storage. Only by making the repositories private and carefully managing user access and user awareness of data security practices can unwanted access be prevented. The Gitlab instance requires two-factor authentication which improves confidence in the identity of users. However, the online Internet-accessible nature of hosted Git repositories carries intrinsic risks.

The servers hosting Git repositories as well as the file servers hosting the network filesystems are located in secure institutional server rooms that allow physical access only for authorized personnel.

Data collected within the frame of LAMASUS will be stored in accordance with partners' institutional requirements that strictly adhere to EU legislation (the GDPR (EU) 2016/679) for data protection and storage. Data are stored in in- house and cloud servers, with back-ups as appropriate. The most popular means of access restriction that will be implemented are:

- password protection
- two-factor authentication
- local network access

6.1. PERSONAL DATA

Specific measures will be taken to obtain informed consent, protect personal data (e.g., encryption, authentication and authorization, and data erasure), minimize the types of data collected and avoid hurting or stigmatizing vulnerable groups or individuals. Data subjects' personal information (e.g. from project partners or stakeholders) will not be made public without their consent.

These contact data collection and processing procedures will comply with national and EU GDPR. All information containing personal data will be stored in a secure manner by consortium partners. Personal data are protected against accidental or lawful destruction, loss, alteration, and disclosure, particularly when processing involves data transmission over networks.

We confirm that these procedures comply with the applicable Austrian, German, and European legal framework, including the General Data Protection Regulation (Regulation (EU) 2016/679) and the national Data Protection Act (1050/2018).

7. Ethics

The most relevant ethical risks for the LAMASUS project is the mis-use of stakeholder information. General ethical issues of informed consent, anonymity and confidentiality associated with the voluntary involvement of human participants from the European Union or its Member States, Associated Countries (Norway, Switzerland), and stakeholders. Information on project members and stakeholders will be stored in a relational database. This will include personal information such as name, e-mail address, institution, country, and some categorical fields related to the role within the project scope. Workshops and interviews with stakeholders will be conducted throughout the project, but we do not foresee any personal data to be recorded.

Specific measures will be taken to obtain informed consent in the collection and processing of data, to safeguard the rights of data subjects under the GDPR, and results will be anonymized when reporting on them. When needed, researchers will be obliged to seek formal ethical approval for the work from the ethics committee of their respective organization, which will comply with all ethical requirements of the country or countries.

LAMASUS portal participation will comply with IT and GDPR rules and guidelines.

8. Concluding remarks

The Data Management Plan has described how the LAMASUS project will manage the generation, collection and processing of environmental data, personal data, and other types of data during and after the project. The DMP is considered a living document and is subject to updates whenever significant changes arise and over time with periodic project evaluations. The purpose of the DMP is ultimately to ensure data under the LAMASUS project is appropriately managed, by providing relevant instruction and guidance to project partners on how proper data management is achieved.

References

Riley, J., 2017. Understanding metadata. Washington DC, United States: National Information Standards Organization (http://www. niso. org/publications/press/UnderstandingMetadata. pdf), 23, pp.7-10.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), pp.1-9.