Transboundary Storm Risk and Impact Assessment in Alpine Regions

RESEARCH DATA MANAGEMENT PLAN

Deliverable 4.1 (D11)

REVISION n.: [01]	DATE: [29/03/2021]	
DISSEMINATION LEVEL: [Public]	WP: 4	TASK(s): 4.1
AUTHORS:	Massimiliano Pittore, Liise Lehtsalu, Silvia Cocuccioni, Kathrin Renner	

Project duration: January 1st 2021 – December 31st 2022 (24 Months)

Main changes compared to previous version	Page(s)

TABLE OF CONTENTS

1 INTRODUCTION

The purpose of this document is to outline the Research Data Management Plan of the project TRANS-ALP. In the next chapter a brief abstract of the project is provided, in order for the document to be self-consistent. In the subsequent section the Data Management Plan is provided.

The Data Management Plan is providing the basic guideline to Data creation / collection, processing and maintenance, and is going to be maintained and updated throughout the project.

2 PROJECT ABSTRACT

The overall goal of the TRANS-ALP project is to develop an integrated methodology for multi-hazard risk assessment and impact forecasting for mountainous regions at cross-border scale, in order to further advance the decision-making processes and support transnational cooperation along the disaster management cycle in the EU. Severe Storms will be addressed as representative examples of complex hazards with cascaded/compounded components (such as floods and landslides) for which a lack of risk assessment and operational protocols has been observed in Alpine regions. Test sites in Italy and Austria will be considered for the exemplification and testing of innovative risk mapping and communication of emerging meteorological threats in the Alps.

3 DATA MANAGEMENT PLAN

3.1 DATA COLLECTION

1.1.1 TYPE OF DATA TO BE COLLECTED OR CREATED

In the project three main types of data are foreseen:

1. Data to be *collected and organised*.

2. Data to be *aggregated or newly created* through processing of heterogeneous sources.

3. Data to be *collected in the field.*

To the **category 1** pertain those data which are already existing but not in a single repository, are not coming from a single source or are not available in comparable formats. Examples of this type of data are:

● Meteorological data (observations, models). These data are usually provided in form of single- and multitemporal gridded (raster) values, according to international open standards, e.g., NetCDF / HDF5. The data size for each event can range from a few MB to several GB in case of extensive scientific simulations.

● Impact and damage data (including geospatial information, narrative text, images and infographics). Data formats in this case include both geospatial formats for raster (e.g.,

GeoTiff) and vector data (e.g., Shapefile) and common "office" documents
(.docx, .xlsx, .csv) used to present data in tabular form and to provide mixed descriptions of
events with text, images, graphs, etc. The expected data size for each event is in the range
10 to 100 MB.

- Large-area geoinformation derived from remote sensing (e.g., forest types, disturbances). These data are usually provided in terms of large georeferenced raster files in standard formats as Geotiff, ASC, etc. expected data size range from 0.1 to 10 GB for each event.

Category 2 comprises, e.g., geospatial data on exposure and vulnerability, based on the harmonisation and integration of different information (e.g., roads, buildings, etc.) that can be further augmented by static and dynamic sources or modified according to the modelling approach. These data will be mostly represented by vectorial geospatial information and described in terms of standard open formats such as geopackage (or shapefile). Expected data size is in the range 1-10 GB.

Category 3 refers to data which are specifically collected within the project to support and integrate the planned activities. For instance:

- in-situ observations and measurements (biomass, snow depth, soil properties, etc.). These data can be provided in several different formats, also including proprietary formats related to specific measurement devices. The expected data size is in the order of 0.1 to 10 GB.

- aerial observations (e.g., high-resolution images, Lidar scanning from manned and unmanned aerial vehicles). Also, these data can be provided in several different formats, also including proprietary formats, in particular related to lidar scanning, whose output is a cloud of 3-dimensional measurements. Data size in this case can be significant, in the range 0.1 to 10 TB.

1.1.2 HOW WILL THE DATA BE COLLECTED OR CREATED

Depending on the typology of data (three categories) different approaches are foreseen. In particular:

Category 1
Data will be provided by different sources, including local and national civil protection agencies and other public bodies (hydromet offices, forest departments), as well as health institutions and other providers. Information and data could be in several languages (Italian, German, English). The following steps will be followed:

1. Definition of one or more *taxonomies* for the categorical organisation of data items. This provides support for both the further organisation of datasets and information and supports the definition of a common naming convention. Where possible and feasible, standard classifications, guidelines or taxonomies will be employed, as for instance the *INSPIRE* Natural Risk Zones Data Specification[1].

2. Collection of data and information from the considered sources. Each information item will be assigned at least the following information: timestamp, location, category, source. Time

---

[1] https://inspire.ec.europa.eu/Themes/140/2892

information will follow INSPIRE metadata specifications, location will include geographical coordinates, where available, and also administrative location in terms of *NUTS* (Nomenclature of Territorial Units for Statistics) and *LAU* (Local Administrative Units).

3. Pre-processing and harmonisation. Data with geospatial attributes (coordinates, areas, places or other toponyms) will be also described using geospatial format (e.g., geopackage). A first Quality Control (QC) will ensure consistency with the objective of data collection and will help setting up the related metadata.

4. Storing and organisation of the collected information. The organisation will be carried out following the LATCH approach (Location, Alphabet, Time, Category and Hierarchy) to provide different means of finding, accessing and evaluating the data. The storing of geospatial data available under an open license will also include the use of geospatial databases *(postgreSQL/postGIS)* and repositories (e.g. *Geonetwork*, *Geonode*) owned and maintained by EURAC Research in order to ease up search and access.

Category 2

Data in category 2 will be mostly based upon data from category 1 plus other data freely accessible or provided by project´s end-users and stakeholders. The following steps will be followed:

1. steps from 1 to 4 as for category 1. Furthermore, in the data collection step specific agreements for data access, use and dissemination will be agreed upon with the providers for the data not already available under open licenses.

2. Processing- this might include different approaches and will lead to an updated, modified or upgraded version of the same datasets, as well as a set of new datasets and models. The specific processing steps and parameters will be described in a technical annex to be linked to the metadata.

3. The newly created data will be assigned a category and the necessary metadata and will be stored accordingly (see step 4 of category 1, above).

Category 3

Data in category 3 will entail the following additional steps:

1. definition of scope, objective and criteria for data collection.

2. design and plan of data collection activities, including any necessary ethics clearance.

3. in-situ data collection / measurement.

4. steps from 2 to 4 as for category 1 will be carried out.

3.2 DOCUMENTATION AND METADATA

Three types of documentation will be provided:

● Official Project´s deliverables

- Additional technical annexes

- Metadata

The project deliverables will be the official reference to the data and provide the operational context as well as the most important summary information and technical specifications.
Where necessary the project deliverables will be integrated by specific technical annexes providing additional technical details, e.g. on the processing methodologies.
Each data item will be assigned a metadata including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Metadata will be provided in text or HTML form according to the most recognised standards, including, e.g., Atom, DIF, Dublin Core, ebRIM, FGDC, ISO[2].

3.3 ETHICS AND LEGAL COMPLIANCE

In the project, ethical issues, if arising, will be addressed according to the following criteria:

- If specific consent will be needed to collect, access, share or preserve information this will be obtained by the relevant parties and appropriately documented (e.g. by signed permission, letter of support, memorandum of understanding, data access agreement). In case of personal data collected from individuals, a suitable information consent procedure will be applied.

- If personal data will be collected, potentially leading to the identification of individuals (this can include observance or surveillance data of a location/territory), it will forego an anonymisation process. For sensitive geospatial data a selective spatial aggregation will be carried out, if necessary, to prevent identification by localization.

- personal data prior to processing (as above) will be transferred and stored securely through encrypted and password-protected formats. It will be then anonymized within the shortest time possible, and the original data will be subsequently deleted. If personal data is collected by one partner but processed by another partner, appropriate processing agreements between Data Controller and Data Processor[3] will be put in place.

3.4 COPYRIGHT AND INTELLECTUAL PROPERTY RIGHTS (IPR)

The data belonging to category 1 will retain the original IPR and copyright according to the provider. Information on IPR / copyright will be stored in the metadata, and access and dissemination will be determined in accordance with the data access agreements.
As for the category 2 data, which entails further processing, a check of consistency with use and re-use licenses will be carried out. For modified and processed data, IPR and licensing will be defined according to the base licenses and the type of processing. If not otherwise specified, new

---

[2] https://en.wikipedia.org/wiki/Metadata
[3] See glossary.

IPR owner will be the one in charge for processing and modification. If necessary, a suitable re-use license will be created. In general, FAIR and open access to data and information will be the guiding principle unless restrictions due to original data licensing apply.

As for the category 3 data, IPR ownership will be retained by those partners responsible or co-responsible for data collection. License of use of collected data will be anyway granted to all project partners within the scope of planned activities.

The specific details of IPR sharing and management will be in accordance with the provisions of the Consortium Agreement.

3.5 STORAGE AND BACKUP

All project data directly related to the implementation of the project activities will be stored within the EURAC digital infrastructure, either as documents, binary file or geospatial dataset. A cloud-based storage solution will be identified, integrated by back-up on high-latency storage devices or through other suitable means. For particularly bulky datasets (e.g., lidar scans) which will be only used within specific activities by some of the partners, alternative storage might be found within the infrastructures managed by other partners. Within the course of the project some of the data will be also stored on openly accessible repositories, either for long-term accessibility (see related section below) or as part of data / paper publication process.

When storing any personal data, storage will be done by Data Controller, if possible. Appropriate GDPR (EU General Data Protection Regulation) compliant solutions in this case will be sought. In case a third-party service will be selected, it will be ensured that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.

1.1.3 ACCESS AND SECURITY

It is estimated that the activities in the TRANS-ALP project are not exposed to specific or relevant risks related to data security. Anyway, several measures to ensure protection of the data while allowing consistent sharing will be put in place. In particular:

- all data repositories including data not explicitly open to general public will be subjected to user authentication, with different levels of protection according to the type of data.

- Sensitive data will be anyway kept under password-protected formats and folders, with a stricter security level with respect to non-sensitive data. These data will also be back upped and transferred using encrypted, password protected formats. If personal data is collected by one partner but processed by another partner, appropriate processing agreements between Data Controller and Data Processer will be put in place.

- Personal data prior to processing will be transferred and stored securely through encrypted and password-protected formats. It will be then anonymized within the shortest time possible, and the original data will be subsequently deleted.

1.1.4 SELECTION AND PRESERVATION

All aggregated, produced or collected datasets (categories 1,2,3) will be immediately (i.e., at creation stage) scrutinized and in case flagged for long-term preservation. Criteria for long-term preservation will, e.g., include:

- Scientific value and potential for further use

- Expected obsolescence / validity interval

- Support and reference to scientific and technical publications

1.1.5 LONG-TERM PRESERVATION PLAN

Data flagged with long-term and / or scientific value will be preserved by publication in international repositories with long-term preservation features, e.g. Pangaea or Zenodo. This will also ensure a straightforward citation, reuse and linkage through the associated DOI (or other permanent identifier). The storing of datasets in long-term repository will be carried out during the project according to the planned activities. A copy of the data stored in long-term repositories will be kept in the partners´ facilities or in EURAC storage solution if necessary for further data processing. All necessary preparation (e.g. update of metadata, use of convenient file format) for long-term preservation will be carried out by the partners responsible of the data themselves, or shared among the consortium.

3.6 DATA SHARING

Data will be shared through different channels:

- data to be used within the consortium will be shared through cloud-based sharing solutions provided by EURAC Research (Geonode / Geonetwork)

- geospatial data will be shared through suitable repositories (e.g. Geonode, Geonework) where data can be stored or linked along with the related metadata.

- whenever possible (see section on long-term preservation, above), datasets will also be published in internationally recognised publishing repositories (e.g. Pangaea, Zenodo) where medatadata can also be automatically harvested by EU-relevant repositories (e.g., European Open Data Portal - data.europa.eu)

1.1.6 RESTRICTIONS ON DATA SHARING

Restrictions to the usage and sharing of data could be related to the following cases (with related contingency measures):

- restrictions due to specific licensing from 3rd party data providers - specific agreements will be set up to grant sharing and use within the consortium

- restrictions due to licensing from project partners - specific agreements will be set up to grant sharing and use within the consortium

- restrictions due to temporary embargo linked to scientific publication - sharing will be pursued to allow for the planned activities, upon agreement that partners will avoid overlapping publications.

- restrictions due to sensitive data - suitable anonymisation and spatial aggregation will be applied to ensure sharing and use of sensitive data.

1.1.7 RESPONSIBILITIES AND RESOURCES

Eurac Research is responsible for implementing the DMP and ensuring it is reviewed and revised. Each project partner will be responsible for data management for the activities he/she is directly responsible (eg, for data of category 3).
For data arising from joined activities, the partner in charge for the specific WP will be responsible of implementing the DMP, and in particular: data capture, metadata production, data quality, storage and backup, data archiving & data sharing. The activities will be anyway carried out in collaboration with EURAC.

1.1.8 RESOURCES TO DELIVER THE DMP

Eurac Research has already the necessary expertise and infrastructure to ensure a proper and comprehensive data management.

3.7 GLOSSARY

Data Controller. Is a legal or natural person, an agency, a public authority, or any other body who, alone or when joined with others, determines the purposes of any personal data and the means of processing it.

Data Processor. Is a legal or a natural person, agency, public authority, or any other body who processes personal data on behalf of a data controller.

Metadata. Metadata is "data that provides information about other data", i.e., it is "data about data". Many distinct types of metadata exist, including *descriptive* metadata, *structural* metadata, *administrative* metadata, *reference* metadata, *statistical* metadata and *legal* metadata.

- Descriptive metadata is descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.

- Structural metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.[4]

- Administrative metadata is information to help manage a resource, like resource type, permissions, and when and how it was created.[5]

- Reference metadata is information about the contents and quality of statistical data.

- Statistical metadata, also called process data, may describe processes that collect, process, or produce statistical data.[6]

- Legal metadata provides information about the creator, copyright holder, and public licensing, if provided.