Project Number: 101121135
Call: HORIZON-CL3-2022-DRS-01-05

Deliverable 7.1

Data management plan

WP 7: Maximizing impact: Dissemination, Communication, Data Management and Stakeholders engagement

Dissemination Level: **PU Public**
Delivery date: **30 April 2024**
Month: **M6**
Organisation leading this deliverable: **ECMWF**

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Find us at: http://gobeyond-project.eu
Twitter: **@GOBEYONDEU**
Linkedin: **gobeyondeu**

Document information

Grand Agreement No.		101121135	Acronym	GOBEYOND
Project Title	GeO and weather multi-risk impact Based Early warning and response systems supporting rapid deploYment of first respONders in EU and beyonD			
Call	HORIZON-CL3-2022-DRS-01			
Topic	HORIZON-CL3-2022-DRS-01-05		Type of Action	Innovation Action
Start Date	01/10/2023		Duration	48 months
Work Package	WP 7: Maximizing impact: Dissemination, Communication, Data Management and Stakeholders engagement			
Deliverable	D7.1: Data Management Plan			
Date of Delivery	Planned	30/04/2024	Actual	02/05/2024
Nature	R — Document, report		Dissemination Level	Public (PU)
Lead Author	Fredrik Wetterhall (ECMWF)			
Email	fredrik.wetterhall@ecmwf.int			
Contributors	Fredrik Wetterhall (ECMWF), Xavi Llort (HYDS), Nikolaos Kekatos (DRAXIS) and Carolina Vega-Viviescas (UPC).			

Document history

Issue Date	Revision	Prepared by	Approved by
22/03/2024	Draft	ECMWF	Fredrik Wetterhall
19/04/2024	Rev-1	Xavi Llort and Nikolaos Kekatos	Fredrik Wetterhall
01/05/2024	Rev-2	Carolina Vega-Viviescas	Daniel Sempere-Torres

Disclaimer

Copywrite message

Executive summary

This document is the Deliverable 7.1 and presents the first version of the **GOBEYOND** Data Management Plan (DMP). This deliverable will be continuously updated over the duration of the project and intermediate versions will be presented at interim reports on M18 and M36, and a final version will be presented at M48.

The updated versions of the DMP will follow each reporting period and will be updated by the coordinator and reviewed by the executive board. The DMP describes the overall strategy for managing the data collected, used, modified and generated during the project. The DMP will be a living document aiming to make data findable, accessible, interoperable and reusable according to the FAIR principle. This will be supported with implementation according to the relevant OGC/ISO standards. In addition, the DMP will provide information on the measures to be taken to guarantee the protection, privacy and security of (sensitive) data during the entire lifecycle of the project.

This document also provides a list of some of the data that are planned to be used in the project. We will however not provide a full list of data used in the project in this deliverable, but rather reference those deliverables. This deliverable is primarily targeted at the consortium partners and should serve as a reference for the management of data products used in the project. It also serves to support the development of new data products in the project. All **GOBEYOND** partners should refer to this plan to ensure a consistent approach to the sharing and documentation of the various datasets (including metadata).

Table of contents

1. Data Summary **1**
1.1. Data collection and production 1
1.2. Data description 2
1.2.1. Hazard data 3
1.2.1.1. Past climate datasets 3
1.2.1.2. Weather Forecast data 3
1.2.1.3. Hydrological datasets 4
1.2.1.4. Forest fire datasets 4
1.2.1.5. Drought datasets 5
1.2.1.6. Other weather datasets 5
1.2.2. Exposure data 5
1.2.2.1. Social exposure 5
1.2.2.2. Physical exposure 6
1.2.3. Vulnerability data 6
1.2.3.1. Social vulnerability 6
1.2.3.2. Physical vulnerability 7
1.3. Purpose of the data beyond GOBEYOND project 7
2. FAIR data **8**
2.1. Making data findable 8
2.2. Making data accessible 8
2.2.1. Repository 8
2.2.2. Data 9
2.2.3. Metadata 9
2.3. Making data interoperable 10
2.4. Increase data re-use 10
2.5. INSPIRE principles 11
3. Other research outputs **12**
4. Allocation of resources **13**
5. Data security **14**
6. Ethical and legal aspects **14**
7. Conclusions **15**
8. References **16**
Appendix I: INSPIRE/ISO requirements on metadata **18**

List of tables

Table 1: The required and recommended metadata that needs to accompany the datasets in **GOBEYOND** 18

List of acronyms and abbreviations

API	Application Program Interface
CADS	Copernicus Climate and Atmospheric Data Store
CDS	Climate Data Store
CERRA (-Land)	Copernicus European Regional ReAnalysis (over land)
CF	Climate and Forecasting
CISI	Critical Infrastructure Spatial Index
CMIP	Coupled Model Intercomparison Project
CLC	CORINE Land Cover
DMP	Data Management plan
DRS	Disaster Recovery (Relief) Systems
DSS	Decision Support System
EC	European Commission
ECA&D	European Climate Assessment & Dataset
ECMWF	European Centre for Medium-Range Weather Forecasts
EDO	European drought observatory
EFAS	The European Flood Awareness System
EFFIS	European Forest Fire Information System
ERA5	Global atmospheric reanalysis from ECMWF (fifth version)
ERA5	HEAT - Universal thermal climate index derived from ERA5
ERA5-Land	ERA5 over land areas
EU	European Union
EW4All	Early warning for all: UN's plan to ensure access to early warning systems for everyone on Earth
FAIR	Findable, accessible, interoperable and reusable
FWI	Fire Weather Index
GDO	Global Drought Observatory
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GEFF	Global ECMWF Fire Forecasting model
GeoTIFF	Georeferences Tag Image File Format
GHSL	Global Human Settlement Layer
GIS	Geographical Information System
GloFAS	Global Flood Awareness System
GPW	Gridded Population of the World

GWIS	Global Wildfire Information System
IFS	Integrated Forecast System
INSPIRE	Infrastructure for Spatial Information in Europe
IoT	Internet of Things
IPR	Intellectual Property Rights
JRC	Joint Research Centre of the European Commission
NetCDF	Network Common Data Form
NUTS	Nomenclature of Territorial Units for Statistics
OGC	Open Geospatial Consortium
OSM	Open Street Map
RDH	The JRC Risk data hub
SoVI	Social Vulnerability Index
SSP	Shared Socioeconomic Pathways
UERRA	Uncertainties in Ensembles of Regional ReAnalysis
UPC	Universitat Politècnica de Catalunya
WMO	World Meteorological Organisation

1. Data Summary

1.1. Data collection and production

In recent years, the amount of available data for risk and impact-based forecasts has been increasing exponentially. With the increase in available sources, such as satellite observations, model simulations, machine learning applications outputs, Internet of Things (IoT), came the heterogeneity in available formats and tools to process the data. One of the tasks of the project is to create an inventory of available data.

The novelty of this project is to combine the data and products from the weather and geo hazards which will very likely lead to difficulties in harmonizing data collection in terms of format, availability and metadata. Therefore, one of the first task of the projects is to map out existing early warning systems (D1.1), the relevant methodologies for enhancing situational awareness (D1.2) and recommendations how to integrate these in **GOBEYOND** platforms (D1.3). Since these are produced in parallel to the Data management plan (DMP), the first full list of available data will be made available after the first interim report at M18.

We are aiming to produce the minimum amount of new data but to re-use the already vast amount of available data sets in existing system. The data used and produced will be on different geographical levels: local, regional and pan-European. The local data will be used mainly in the local decision support systems (DSS), and the Mediterranean (EW4MED) and pan-European (EW4Europe) early warning platforms. The main data to be used for the hazard are geo and weather monitoring systems, and numerical weather prediction NWP for the weather hazard forecasts, and reanalysis, and observational data for the One of the main sources of data for this will be the Climate Data Store (CDS). See below a description of some of the datasets in CDS. The vulnerability and exposure data will be either local data or taken from free open data sources like the Joint Research Centre portal and Risk Data Hub.

The Climate Data Store (CDS)[1] provides access to a wide variety of climate-related data, including observational data from ground-based instruments, satellites and other sources, reanalysis data, and environmental and climate model projections.

In addition to these main types of data, the CDS also provides access to a range of derived products, such as indices of extreme weather events, indicators of climate change impacts, and data processing and visualization tools available through the CDS Toolbox. The toolbox is currently under redevelopment and a new version will soon be available with much more flexibility for the user to incorporate the code into local applications since it will be available on GitHub. The data can be accessed in various formats, such as netCDF, GRIB, and CSV, and can be used for a range of applications, including research, policy-making, and commercial activities.

Data in CDS follow strict quality control and guidance when it comes to the FAIR principle and metadata. It is fully INSPIRE[2] compliant. The data produced within the **GOBEYOND** project will most likely not always comply with the strict requirements of CDS. For those datasets the CMP will provide a guideline on what metadata is needed. Data in the CDS

[1] https://cds.climate.copernicus.eu/#!/home
[2] https://knowledge-base.inspire.ec.europa.eu/index_en

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

is programmatically available through the Climate Data Store (CDS) Application Program Interface (API) service and CDS Toolbox.

The Joint Research Centre's (JRC) risk data hub[3] provides access to a range of data related to natural hazards and environmental change, with a particular focus on Europe. The data can be broadly categorized into three main types:

- Hazard data: The datasets that describe the physical characteristics and potential impacts of natural hazards, such as floods, landslides, wildfires, droughts, and coastal erosion. Some of the hazard data available in the hub include flood maps, landslide inventories, and drought indices.
- Exposure and vulnerability data: These are datasets that describe the elements at risk, such as population, buildings, infrastructure, and ecosystems, as well as their susceptibility to damage or disruption from natural hazards. Examples of exposure and vulnerability data available in the hub include population density maps, building footprint datasets, and land-use classifications.
- Risk data: These are datasets that combine hazard, exposure, and vulnerability data to estimate the likelihood and potential consequences of natural hazards in a given area. Examples of risk data available in the hub include flood risk maps, earthquake loss models, and multi-hazard risk assessments.

The data in the hub are collected from various sources, including satellite imagery, ground-based measurements, and modelling tools. The data are typically available formats that are compatible with Geographical Information systems (GIS), such as shapefiles, GeoTIFFs, and netCDFs, and can be used for various applications, including risk assessment, disaster management, and climate adaptation planning.

When we identify that the data is needed but not available yet and some partners can generate the data, it will be deposited in the well-established data repositories. The project leader board will during the first 18 months of the project investigate which repository is best fit for the purpose (see section 2.2). This includes, but is not limited, to population projections and downscaling of climate data to higher resolution to be useful for regional climate risk assessment. If the data produced in the **GOBEYOND** project, proves to be useful for the communities beyond the project, it will be made available through other sources such as CDS or JRC RDH for long-term preservation after the project ends if possible due to license issues.

1.2. Data description

Data plays a vital part in the **GOBEYOND** impact forecasts. The project aims to collect and generate environmental simulations and observation data to produce forecasts of impacts and risks, including hazards, vulnerability, and exposure to weather and geo hazards. The data will be used to aid the assessment of risk and potential impacts of extreme weather and geo on society and the environment. This will be useful material for developing strategies to mitigate and adapt to these risks by regional agencies in European Union (EU).

Main datasets identified at the beginning of the project are on local, regional and European scale, available for all **GOBEYOND** partners. Work package (WP) 1 has the tasks to identify and describe these datasets and available methodology and forecast systems.

[3] https://drmkc.jrc.ec.europa.eu/risk-data-hub#/

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

They are divided into three groups: hazard, vulnerability and exposure data. All datasets described below are open source.

1.2.1. Hazard data

Hazard data used should as far as possible be data with the highest resolution to reflect the local environment. However, it is not always possible to retrieve those data, therefore as a minimum, we will always be able to provide generic and hazard data from the European and global data sets available in the Copernicus Climate Data Store (CDS, see below for a description). The CDS data is following all the principles of FAIR so it guarantees the quality of the data.

1.2.1.1. Past climate datasets

E-OBS gridded observations is an ensemble dataset available at 0.1 and 0.25 degree regular grid for daily mean, minimum and maximum temperature, daily precipitation daily mean sea level pressure, daily mean wind speed, daily mean relative humidity and global radiation QQ. The dataset covers the European domain: 25N-71.5N x 25W-45E. The data files are in NetCDF-4 format. The dataset is calculated from station data acquired through the network of the European Climate Assessment & Dataset (ECA&D) project.

ERA5 and ERA5-Land reanalysis are the global atmospheric reanalysis datasets produced by the ECMWF. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. The datasets provide a comprehensive view of the Earth's atmosphere, for the period 1940 to present for ERA5 (resolution 0.25 degrees) and 1959 to present for ERA5-Land (0.1 degree resolution). Both reanalysis data sets are widely used in climate research, weather forecasting, and other applications. ERA5 also includes an ensemble component at half the resolution to provide information on the synoptic uncertainty of its products. ERA5-land parameter uncertainty currently can be accessed using the equivalent ERA5 fields.

The **UERRA** (Uncertainties in Ensembles of Regional ReAnalysis) data set contains reanalysis data of the atmosphere, the surface and near-surface as well as for the soil covering Europe. Essential climate variables are generated with the UERRA-HARMONIE and the MESCAN-SURFEX systems. UERRA uses a combination of advanced data assimilation techniques and high-resolution regional climate models to produce a climatology of the atmosphere over Europe at a resolution of 0.1 degrees, covering the period 1961-2019. In addition to the standard atmospheric variables, UERRA also provides uncertainty estimates for each variable, which can be used to assess the reliability of the data and quantify the level of confidence in any climate-related analysis or decision-making process.

The **CERRA and CERRA-Land** (Copernicus European Regional ReAnalysis Land) dataset provides regional reanalysis over the European domain on a Lambert conformal conical grid of 5.5 x 5.5 km over the period September 1984 to April 2021. The parameters contain atmospheric and land variables, with soil parameters provided on different soil depths.

1.2.1.2. Weather Forecast data

The available forecast data will come from several numerical weather prediction data, nowcasting and direct monitoring. The data will depend on the scale and specificity of the application. As an example, for the European case the ECMWF Integrated Forecast

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

System (IFS) ensemble will be sued, consisting of 51 ensemble members run twice daily (4/day with smaller ensemble size) with a spatial resolution of 9 km. For the EWMED and other platforms, the used data will be specified in upcoming versions of the DMP.

1.2.1.3. Hydrological datasets

The European Flood Awareness System (EFAS), jointly developed by the European Commission and the ECMWF, is a hydrological forecast and monitoring system covering the greater European domain. This dataset was produced by forcing the LISFLOOD hydrological model with gridded observational data of precipitation and temperature at a 5x5 km resolution across the EFAS domain. The historical data is available from 1991-01-01 up until near-real time, at the 6 hourly time step, with a delay of 6 days. The hydrological forecasts are available from 2018-10-10 up until the present with a 30-day embargo. This embargo has been decided by the EFAS partners to protect the one-voice principal when it comes to issuing flood warnings from official sources. The data will be made available to partners within the consortium but cannot be shared to the public.

Global hydrological data sets produced by the Global Flood Awareness System (GloFAS) includes river discharge and related historical, forecast and forecast data, as well as seasonal forecast for users looking for longer lead-time forecasts. They offer a consistent representation of a key hydrological variable across the global domain at 0.1° x 0.1° horizontal resolution. Historical data covers the period from 1979 to near real-time, while the forecast is available from 2019 and updated daily for 30 days forecast with a daily step.

GLoFAS is a collaborative effort between the European Commission's Joint Research Centre and the European Centre for Medium-Range Weather Forecasts to provide early flood warnings globally by combining weather forecasts with hydrological models. It is the “Global” version of EFAS and will provide (limited) data beyond Europe to cover project needs in the African pilot.

1.2.1.4. Forest fire datasets

Fire danger indices historical data from the Copernicus Emergency Management Service data set provides a historical reconstruction of meteorological conditions favourable to the start, spread and sustainability of fires. The fire danger metrics provided are part of a vast dataset produced by the Copernicus Emergency Management Service for the European Forest Fire Information System (EFFIS). The European Forest Fire Information System incorporates the fire danger indices for three models developed in Canada, the United States and Australia. In this dataset, the fire danger indices are calculated using weather forecasts from historical simulations provided by ECMWF ERA5 reanalysis. Data is available at 0.25° x 0.25° horizontal resolution, with a daily step for the period 1979 to the present.

GWIS (Global Wildfire Information System) is a platform developed by the European Commission's Joint Research Centre to provide global wildfire monitoring and information exchange for better wildfire management and response. GWIS will provide fire danger information beyond Europe in the project activities

Fire danger indicators for Europe from 1970 to 2098 derived from climate projections data set present projections of fire danger indicators for Europe based upon the Canadian Fire Weather Index System (FWI) under future climate conditions. The FWI is a meteorologically based index used worldwide to estimate fire danger and is implemented

This project has received funding from the European Union’s Horizon Europe programme for research and innovation under grant agreement No. 101121135.

in the Global ECMWF Fire Forecasting model (GEFF). In this dataset, daily FWI values, seasonal FWI values, and other FWI-derived, threshold-specific, indicators were modelled using the GEFF model to estimate the fire danger in future climate scenarios. The dataset is available at NUTS2 regions.

1.2.1.5. *Drought datasets*

The European Drought observatory of the JRC provides a number of drought datasets, both for monitoring and forecasting purposes. These data are available to download through the European drought observatory (EDO) web portal, but we will also most likely create other drought indices using available monitoring and forecast data within the platforms.

The Global Drought Observatory (GDO) is an initiative led by the European Commission's Joint Research Centre to provide timely and accurate information on drought conditions worldwide through satellite-based monitoring and analysis. GDO will extend beyond Europe the EDO role.

1.2.1.6. *Other weather datasets*

The CDS also contains various other datasets for the global and European domain, such as the universal thermal climate index derived from ERA data (ERA5-HEAT), windstorm data such as the synthetic windstorm events in Europe and seas water level changes from today and to the future. CDS contains most of the climate and hazard data that are potentially useful in a Climate Risk Assessment, however, they are on a quiet coarse resolution for a very detailed analysis. We will during the first phase of the project further investigate which datasets will be used in **GOBEYOND**.

1.2.2. Exposure data

The Europe-wide exposure data to be used in **GOBEYOND** will primarily be open-access data selected based on the data needs in terms of spatial and temporal resolution and underlying acquisition or modelling approach.

1.2.2.1. *Social exposure*

To characterize social exposure, a range of population datasets are available, with different spatial and temporal resolutions, such as the satellite-based Global Human Settlement Layer (GHSL) population data GHS-POP, available at 100 m/3 arcseconds and 1 km/30 arc seconds from 1975 to 2030 in 5-year time steps (European Commission & Joint Research Centre, 2023). Further, WorldPop provides population data in annual time steps for the period 2000-2020 based on a Random Forest modelling approach (Stevens et al., 2015). GHS-POP and WorldPop are based on the Gridded Population of the World (GPW) (version 4) that have a spatial resolution of 30 arc seconds and a temporal resolution of 5-year time steps from 2000-2020 (CIESIN - Center for International Earth Science Information Network - Columbia University, 2018b), which were collected from the national census and population registers. On an administrative unit level, i.e. NUTS (Nomenclature of Territorial Units for Statistics) regions, population data are available from Eurostat as well as the Joint Research Centre's (JRC) Risk Data Hub (see below).

Future projections of the population until 2100 are available under a range of socioeconomic scenarios, the Shared Socioeconomic Pathways (SSPs) (O'Neill et al.,

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

2017). Publicly available datasets include those of Merkens et al. (2016), Wang et al. (2022), and Gao (2017), all of which are available at a spatial resolution of 30 arc seconds, however, based on different modelling approaches and underlying population data used as model input. Further population projections are available from GHSL at 1 km spatial resolution, and from Terama et al. (2019) at NUTS 2 level.

1.2.2.2. Physical exposure

Several datasets are available to characterize physical exposure. GHSL provides raster data of built-up land and volume, residential and non-residential settlement zones (Morphological Settlement Zones), settlement classes, and building height at 2 m to 1 km/ 30 arc seconds spatial resolution and a temporal resolution of five-year time steps from 1975 to 2030 for most datasets. Furthermore, built-up land, infrastructure, protected land, and land-use data are available per administrative unit for different European NUTS regions (i.e. NUTS 0-3) as well as in raster format from the JRC Risk Data Hub. Open Street Map (OSM) provides spatial data in vector format (i.e. points, lines, polygons) of e.g. building footprints and types, health and education facilities, energy and telecommunication towers, and roads and railway networks (www.openstreetmap.org). The Critical Infrastructure Spatial Index (CISI) is based on OSM data and is available in raster format at a spatial resolution of 0.25 degrees (Nirandjan et al., 2022). For assessing the exposure of different land uses, the Europe-wide CORINE land cover data are available for 44 land cover classes at 100 m spatial resolution for the years 1990, 2000, 2006, 2012, and 2018 (https://land.copernicus.eu/).

1.2.3. Vulnerability data

Similar to the exposure data, **GOBEYOND** focuses on using publicly available, European-wide datasets to characterize vulnerability.

1.2.3.1. Social vulnerability

For social vulnerability, demographic data are available from GPW, including age and sex raster data at 30 arc seconds spatial resolution for the year 2010 (CIESIN - Center for International Earth Science Information Network - Columbia University, 2018a), as well as for the years 2000-2020 from WorldPop at 3 and 30 arc seconds spatial resolution. The Global Data Lab further provides data on education levels, income and inequality at administrative unit level, roughly corresponding to NUTS 2 level (Smits & Permanyer, 2019). Similarly, the World Bank provides administrative unit-level data on several poverty indicators such as daily consumption and the number and ratio of poor people (Rentschler et al., 2022). The JRC RDH includes several datasets mainly at NUTS 2 level that can be used as indicators of vulnerability such as life expectancy, education levels, household income, and employment.

Currently, hardly any social vulnerability projections datasets are available at European scale, an exception being age and sex structure data at NUTS 2 level for a selected set of SSPs (Terama et al., 2019). Furthermore, several studies produced projections of Gross Domestic Product (GDP) based on SSP-based spatial population projections, e.g. at 0.5 degree (Murakami & Yamagata, 2019) and 1 km (van Huijstee et al., 2018) spatial resolution; however, the latter are currently not publicly available. Vulnerability projections that may be developed as part of **GOBEYOND** will be made publicly available in a well-established data repository.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

1.2.3.2. *Physical vulnerability*

Physical vulnerability is often assessed with the help of damage curves, assuming that structures (e.g. buildings, infrastructure) will suffer a certain damage based on hazard intensity and the structure materials. Therefore, building materials data can be used from the Pager database (Jaiswal et al., 2010). For flood vulnerability, damage curves are available for different buildings and land use types (Huizinga et al., 2017).

The size of the data will vary depending on the dataset, ranging from a few megabytes to several gigabytes per data set for climate projections. We will identify the most important data that will be copied to the **GOBEYOND** data repositories for easier use due to its size or other constraints.

1.3. Purpose of the data beyond GOBEYOND project

Most of the data used in the **GOBEYOND** project will come from well-established data providers. The data from JRC's Risk Data Hub is available for download using the web portal and interface, but not through the API. The existing applications and systems will have their well-established data feeds and repositories.

Moreover, all impact forecasts that are developed in the context of the **GOBEYOND** processing core will be open source and free for use everyone as far as possible and it does not violate the General Data Protection Regulation (GDPR) of the data. The principle "as open as possible, as closed as necessary" will be applied. The methodologies will be published in international journals, and the repositories will offer the data to be used by anyone, given the restrictions attached to each of the datasets.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

2. FAIR data

Most of the tools and data provided by **GOBEYOND** will be based on data from existing portals that are already community or EC resourced and targeted to societal applications such as GEOSS and Copernicus data. Access to these data is provided via standard interfaces and will be provided in formats complying with Open Geospatial Consortium (OGC) standards, Climate and Forecast (CF) standards and regular metadata documentation protocols.

All data generated throughout the course of the project will be made available applying the FAIR principle: data will be Findable (through standards for identification and rich metadata), Accessible (how data are accessed during and after the project lifetime), Interoperable (defining sharing policy and integration into workflows) and Reusable (including clear licensing and correct formatting of data and metadata).

Copyrighting, licensing and not knowing which repository to use for storing data are the highest challenges researchers face when publishing the data (Stuart, et al., 2018). Documentation about standards and examples of managing datasets will be added in the next version of the DMP.

2.1. Making data findable

Making data findable involves ensuring that data is discoverable and identifiable by a persistent identifier, and that rich metadata is provided to allow for easy discovery. In order to achieve this, the data produced in the project will be accompanied by metadata that follows the domain-specific standards, such as the CF convention and OGC standards for hazard data, and the vulnerability and exposure data the standards used in JRC Risk Data Hub. The guidance will be provided to the pilot partners and participants of the open call for their data management.

In cases where metadata standards do not exist in the discipline, the project will outline what type of metadata will be created and how it will be created. Search keywords will also be provided in the metadata to optimize the possibility for discovery and potential reuse. Additionally, the metadata will be offered in a way that can be harvested and indexed. This will enable national and European data portals to index and harvest the metadata, which will make the data more easily discoverable and accessible to a wider audience. Overall, making data findable is crucial for enabling effective data sharing and reuse, which can lead to more impactful research and innovation during and after the project.

2.2. Making data accessible

2.2.1. Repository

To make the data produced by the **GOBEYOND** project more accessible to the project partners, all data collected and generated throughout the project will be stored on secure servers with proper backup procedures in place to prevent data loss or corruption. Data that are directly from the Copernicus Climate and Atmosphere Data Store (CADS) or JRC Risk Data Hub (RDH) will be stored on servers, but not backed up. The scripts that are used to access the data will be saved so that if the data is compromised it can be easily restored. Access to this data will be granted to project partners and stakeholders through

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

secure and efficient means, with data privacy and security ensured by ECMWF's access control management. Unless Pilot partner institutions use local data that cannot be shared, all data generated by the **GOBEYOND** project will be publicly available by default.

The project's data will also be made available to the wider scientific community through open data platforms and repositories that include appropriate metadata and documentation to facilitate data reuse and citation.

For depositing data which accompany research papers, we will investigate different online repositories, including Zenodo and Figshare. Zenodo[4] is OpenAIRE's open-access repository for EC-funded research, providing barrier-free access to research outputs. Meanwhile, [5] allows users to make their research outputs available in a citable, shareable, and discoverable way. Both repositories provide data identifiers and services are available for uploading and downloading the data via open REST API. We will also be able to store data in the European weather cloud (EWC) or our own cloud infrastructure (CCI), which will be the systems we will use for hosting the **GOBEYOND** processing core. The project partners will assess the suitability of each repository for use in the project and will provide a tutorial for those new to the chosen repository. The results of this assessment will be presented in the next version of this data management plan.

Should any of the datasets produced during the project prove to be useful to the wider European community, they can be proposed to become deposited in well-established, domain-specific data repositories, namely CADS for hazard data, and RDH for exposure and vulnerability data.

2.2.2. Data

The project is striving to use and produce open data wherever possible, although certain data may not be openly available due to ownership or other constraints. Local vulnerability and exposure data, as well as possibly local meteorological and geo data, will often be owned by a third party and only available to the project partners and stakeholders and therefore it cannot be opened.

However, the data produced in the project will be when possible be made accessible through a free and standardized access protocol, namely a free and open REST API. The REST API will enable controlled access to impact forecasts. As there is no personal or sensitive data involved, there is no need for a data access committee to evaluate or approve access requests. The project will potentially also use the EFAS forecasts. The forecasts which are less than 30 days old are restricted to only EFAS partners, but they can be used for research purposes of they are not spread.

2.2.3. Metadata

All metadata used in the project will be stored together with the data description, available through a public domain. The metadata in organised repositories such as the Climate data store is well described with each dataset. For the datasets that are collected from other sources we will provide a dedicated space within the project repositories for metadata, with the possibility to enhance the metadata. The data produced as a result of the project will follow OGC standards[6], to be defined and presented in the second version of the DMP.

[4] https://zenodo.org/
[5] https://figshare.com/
[6] https://www.ogc.org/standards/

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

The software used in the project will be all open software, and all libraries will be described within the WP3 alongside references to metadata and licenses.

2.3. Making data interoperable

We aim to make the data produced and used in the **GOBEYOND** project interoperable within and across disciplines. To achieve this, we will follow established community-endorsed best practices, such as using common and widely accepted standards and formats, providing rich metadata, using persistent identifiers, and making the data openly available wherever possible. For hazard data, the project will follow standards defined by the World Meteorological Organisation, CF convention, and OGC standards. For exposure and vulnerability data, we will explore and follow the standards, vocabularies, and methodologies used in JRC RDH, such as standardized terms for describing the physical characteristics of buildings and infrastructure, as well as terms for describing social and economic characteristics. We also plan to use JRC RDH's standardized classification system for land use and land cover. By using these controlled vocabularies, we will make it easier to compare and combine data from different sources.

We will provide mappings to more commonly used ontologies in case we need to generate project-specific ontologies or vocabularies and will also make the generated ontologies or vocabularies openly available for reuse, refining, or extending. In addition, we will include references to the datasets from previous research, to ensure that the data can be easily integrated with other datasets.

We will also make an inventory of all types and formats of data used for impact forecasts for climate risk assessment and provided by the CDS and RDH and make a list of recommendations for the project partners which will be publishing the data as part of their climate risk assessment.

2.4. Increase data re-use

The aim of FAIR data is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated or combined in different settings. This means that (meta)data are richly described with a plurality of accurate and relevant attributes, are released with a clear and accessible data usage license, are associated with detailed provenance, and meet domain-relevant community standards.

The licensing of the **GOBEYOND** tools and data will follow the best practice for free sharing with the public. More details on this will be provided in subsequent versions of the DMP. Some data that will be used by the project are under specific licenses, such as the data in the European climate data store. Other data will come with a specific license to use. In those cases, the project will make sure that all users have the right to use the data and clearly state the principles of reusing and sharing the datasets.

All relevant datasets produced by **GOBEYOND** will be made public, if possible either as deliverables or as part of a data describing article in peer-revied literature. For larger datasets, details will follow in later versions of this DMP how they will be made available. The starting point would be to store this data in a common repository that is accessible for download by individual users

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

2.5. INSPIRE principles

To implement the FAIR principles in the project we will make use of the INSPIRE Directive[7] of the EU. INSPIRE principles are that data should be:

- Collected only once and kept where it can be maintained most effectively.
- Possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- Possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Contain Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

The implementation of the INSPIRE in **GOBEYOND** can be found in the Appendix 1, showing what metadata is needed to collect for each dataset to be fully compliant. The metadata have some data that is mandatory (M) and some that are optional (O). It is recommended to also have the optional metadata, but the data will pass the INSPIRE requirements with also just the mandatory requirements.

The DMP will in the first reporting period be accompanied by a template where the users can provide the metadata for each dataset used and produced within the project.

[7] https://knowledge-base.inspire.ec.europa.eu/overview_en

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

3. Other research outputs

All public deliverables and relevant milestones, training material, journal articles and conference contributions, as well as surveys and questionnaires (without sensitive data) will be made publicly available as early as possible, including as drafts, pre-prints or registered reports. All code development and code documentation within the project that are not proprietary will be shared with users and developers through public git repositories. This allows not only a shared responsibility for developing tools, but also a further uptake of the developments in the wider community.

The internal communication within the project is through SharePoint which is hosted by the coordinator, the Universitat Politècnica de Catalunya (UPC).

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

4. Allocation of resources

The costs required for making the data collected and generated FAIR have been included in the budget of the project. UPC as project coordinator have the overall responsibility for the delivery of the project, the ECMWF and the Task 7.2 team have the primary responsibility for creating, updating and delivering the data management plan. However, all project partners are involved, either in support of the overall data management plan or by ensuring their contributions adhere to the FAIR principles. The related costs have therefore been included in the overall contribution of each partner.

The overall storage of data will primarily be done by the owner of the data, and the processing core will connect to this data when needed. Data that needs to be stored locally will be stored at a cloud solution that is yet to be defined, depending on what the overall technical solution is for the implementation. This will be further defined in the second version of the DMP in M18. A solution for the final storage of the data will be provided in the final version of the DMP which will be delivered in M48 and it is envisioned that this will encompass data being stored in the CDS as well as in the JRC Risk data hub. ECMWF has dedicated 9PM to WP7, and most of this will be dedicated to the implementation and updating of the DMP.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

5. Data security

The data used in the project will not by default have any security. The data used within will be mostly from free and open sources. In the cases where the user will bring data that is sensitive and cannot be shared freely, this data will be protected via authentication on servers that are within the processing core. Project partners can also choose to run the scripts and setup the services within in their own environment to make sure that data is not shared. All internal documents and communications are shared with project partners through SharePoint, which is hosted by the project coordinator.

In terms of reproducibility, the processing core will make all data reproducible, and therefore robust. Any pre-processed data from other sources, such as CDS or ECMWF, will be stored on disk for ease of access, but the data will not be backed up more than necessary. However, all scripts will be backed up to be able to reconstruct any vital data should they be corrupted.

The long-term plan for storage of the data will be within the existing European entities, such as the CDS and the Risk Data Hub, if the data is deemed valuable for a broader audience. We will also explore if the data can be made part of the DestinE data lake. Other data will be stored on disk for the duration of the project, but after that it is the responsibility of the user to store output. However, since the products designed within the project are designed to be reproducible, any output can be reproduced.

6. Ethical and legal aspects

The project will not record or save any personised data from the users of any system. For the purposes of accessing restricted information, the users will be asked to provide a username and an email address which will be maintained on a dedicated server hosted by ECMWF.

All data accessible through login on the project website and data management platform will be stored in conformity with Intellectual Property Rights (IPR), and the new General Data Protection Regulation (GDPR), which require data collection and management to be lawful, fair and transparent.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

7. Conclusions

This first version of the DMP sets out the principles for the proper management of the various datasets used and produced in the project. The next step will consist in collecting more details from **GOBEYOND**'s partners, about the datasets they are, or will be using and producing. The provided information will be checked against the FAIR principles and adjustments will be made, where needed.
There will be regular updates of the DMP and the next version of the DMP will be produced by M18 of the project. This will contain more information on the storage of the data within the project, as well as a larger catalogue of the data used within the project. Intermediate updates of this DMP will be available at all times throughout the project. The final version of the DMP will be delivered at M48, and this will provide a detailed plan for the long-term plan for storage of the output.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

8. References

CIESIN - Center for International Earth Science Information Network - Columbia University. (2018a). Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 11. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45H7D7F

CIESIN - Center for International Earth Science Information Network - Columbia University. (2018b). Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/https://doi.org/10.7927/H4PN93PB

European Commission, & Joint Research Centre. (2023). GHSL data package 2023. Publications Office of the European Union. https://doi.org/doi/10.2760/098587

Gao, J. (2017). Downscaling global spatial population projections from 1/8-degree to 1-km grid cells. NCAR Technical Note NCAR/TN-537+STR, National Center for Atmospheric Researcher, Boulder, CO, USA. https://doi.org/10.5065/D60Z721H M4 - Citavi

Huizinga, J., de Moel, H., & Szewczyk, W. (2017). Global flood depth-damage functions. In J. R. C. European Commission (Ed.), EUR 28552 EN. https://doi.org/10.2760/16510 T4 - Methodology and the database with guidelines T5 - JRC Technical Reports M4 - Citavi

Jaiswal, K., Wald, D., & Porter, K. (2010). A Global Building Inventory for Earthquake Loss Estimation and Risk Management. Earthquake Spectra, 26(3), 731–748. https://doi.org/10.1193/1.3450316

Merkens, J.-L., Reimann, L., Hinkel, J., & Vafeidis, A. T. (2016). Gridded population projections for the coastal zone under the Shared Socioeconomic Pathways. Global and Planetary Change, 145, 57–66. https://doi.org/10.1016/j.gloplacha.2016.08.009

Murakami, D., & Yamagata, Y. (2019). Estimation of Gridded Population and GDP Scenarios with Spatially Explicit Statistical Downscaling. Sustainability, 11(7), 2106. https://doi.org/10.3390/su11072106

Nirandjan, S., Koks, E. E., Ward, P. J., & Aerts, J. C. J. H. (2022). A spatially-explicit harmonized global dataset of critical infrastructure. Scientific Data, 9(1), 1–13. https://doi.org/10.1038/s41597-022-01218-4

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., & Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. Global Environmental Change, 42, 169–180. https://doi.org/10.1016/j.gloenvcha.2015.01.004

Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. Nature Communications, 13(1), 3527. https://doi.org/10.1038/s41467-022-30727-4

Smits, J., & Permanyer, I. (2019). Data descriptor: The subnational human development database. Scientific Data, 6, 1–15. https://doi.org/10.1038/sdata.2019.38

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLOS ONE, 10(2), e0107042. https://doi.org/10.1371/journal.pone.0107042

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., & Astell, M. (2018). Practical challenges for researchers in data sharing. Springer Nature.

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

Terama, E., Clarke, E., Rounsevell, M. D. A., Fronzek, S., & Carter, T. R. (2019). Modelling population structure in the context of urban land use change in Europe. Regional Environmental Change, 19(3), 667–677. https://doi.org/10.1007/s10113-017-1194-5

van Huijstee, J., van Bemmel, B., Bouwman, A., & van Rijn, F. (2018). Towards an Urban Preview: Modelling future urban growth with 2UP.

Wang, X., Meng, X., & Long, Y. (2022). Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. Scientific Data, 9(1), 563. https://doi.org/10.1038/s41597-022-01675-x

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.

Appendix I: INSPIRE/ISO requirements on metadata

Table 1: The required and recommended metadata that needs to accompany the datasets in **GOBEYOND**

Metadata part	Element	Sub-element	ISO	Obligation	Domain	Example
Identification	Title		title	M	Free	
Identification	Abstract		abstract	M	Free	
Identification	Type		hierarchy_level	M	Code_List (B.5.25)	Dataset
Identification	Location		linkage	O	url	Where data can be downloaded
Identification	URI	code+codeSpace	identifier	M		DOIs
Identification	Language		language	O	Code_List (ISO 639-2)	Italian - ita
Classification	Topic and category		topicCategoty	M	Enumeration (B.5.27)	Climatology Meteorology Atmosphere
Keyword	Keyword_value		keyword	M	Free	Controlled Vocabulary
Keyword	Originating_vocabulary		thesaurys_name	O	title+reference_date (dateType, date)	
Geographic Location	BBOX	West, East, South, North	XBound	M	180 90	
Temporal Reference	Temporal	Extent, publication, last revision, creation		M	At least 1 of them	
Quality Validity	Lineage		statement	M	Free	
Quality Validity	Spatial resolution	eqivalent scale, distance		O		50000 0.25
Conformity	Degree		pass	M	Boolean (true. False. Null)	
Conformity	Specification		specification	M	title+reference_date (dateType, date)	
Constraints related to access & use	Limitations on public access	accessConstraints, otherConstraints, classification	accessConstraints	O		

This project has received funding from the European Union's Horizon Europe programme for research and innovation under grant agreement No. 101121135.