Sustainable Historic Environments holistic reconstruction through Technological Enhancement & community-based Resilience

D.1.1. Data sources and Knowledge for CH climate and disaster resilience

Version number	2.0
Dissemination level	PU
Lead partner	SISTEMA
Due date	31.05.2020
Type of deliverable	report
Status	delivered

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821282

Published in the framework of:

SHELTER - Sustainable Historic Environments hoListic reconstruction through Technological Enhancement & community-based Resilience

Authors:

Marco Folegani, Maria Luisa Quarta (SISTEMA, WP1 Leader)

Daniel Marco, Jorge Ruperez (EGIS); Gemma Garcia Blanco, Nieves Peña Cerezo (TECNALIA); Luc Pockele (RED); Edoardo Arnaudo, Alessandro Farasin, Vito Macchia, Claudio Rossi (LINKS); Petr Mikuš, Viliam Zathurecky (UMAS)

Revision and history chart:

VERSION	DATE	AUTHOR/EDITOR	COMMENT
0.1	26/03/2020	SISTEMA	Table of content
0.2	30/04/2020	SISTEMA	First draft to be circulated among contributors
0.3	08/05/2020	SISTEMA, EGIS, TECNALIA, RED, LINKS, UMAS	Second draft with contributions from partners
0.4	18/05/2020	SISTEMA, EGIS, TECNALIA, RED, LINKS, UMAS	Submitted to reviewers
0.5	25/05/2020	SISTEMA	Submitted to PC
1.0	04/06/2020	SISTEMA	Final version
1.1	24/09/2020	SISTEMA	Integration of PO feedback
2.0	21/05/2021	SISTEMA	Integration of EC reviewers feedback (update of reference list)

Disclaimer:

Table of content

List of tables

List of figures

Definition, Acronyms and Abbreviations

Acronym	Full name
AI	Artificial Intelligence
AMS	Asynchronous Messaging System
AoI	Area of Interest
API	Application Programming Interface
ARPAE	Regional Agency for Prevention, Environment and Energy
C3S	Copernicus Climate Change Service
CAMS	Copernicus Atmosphere Monitoring Service
CC	Climate Change
CCA	Climate Change Adaptation
CH	Cultural Heritage
CHA	Corine Land Cover Change
CHM	Cultural Heritage Management
CLC	Corine Land Cover
CLMS	Copernicus Land Monitoring Service
CMEMS	Copernicus Marine Environment Monitoring Service
CSS	Copernicus Security Service
DEM	Digital Elevation Model
DMF	Data Mapping Form
DRD	Data Resilience Dashboard
DRM	Disaster Risk Management
DDP	Data Driven Platform
DMP	Data Management Plan
DRR	Disaster Risk Reduction
DSS	Decision Support System
EC	European Commission
ECHO	European Civil Protection and Humanitarian Action
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EEA	European Environment Agency
EM-DAT	Emergency Events Database
EMS	Copernicus Emergency Management Service
EO	Earth Observation
ESA	European Space Agency
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GCOS	Global Climate Observing System
GIS	Geographic Information System
GML	Geography Markup Language
HA	Historic Area
ICOMOS	International Council on Monuments and Sites
ICT	Information and Communication Technologies
IH	Intangible Heritage
IPR	Intellectual Property Rights
ISRBC	International Sava River basin Commission

ISTAT	Istituto Nazionale di Statistica
KPIs	Key Performance Indicators
MiBACT	Istituto Superiore per la Conservazione ed il Restauro
ML	Machine Learning
NDA	Non-Disclosure Agreement
NDVI	Normalized Difference Vegetation Index
NH	Natural Heritage
OLs	Open Labs
OLTWG	Open Lab Task Working Group
R&I	Research & Innovation
RACER	Relevance, Acceptability, Clarity, Easiness, Robustness
RDA	Rapid Damage Assessment
RFID	Radio Frequency Identification
RM	Rapid Mapping
RRM	Risk and Recovery Mapping
S1	Sentinel 1
S2	Sentinel 2
SAR	Synthetic Aperture Radar
SC	Steering Committee
T	Task
TA	Tangible Heritage
UCS	Use case Scenarios
UI	User Interface
URs	User Requirements
VR/AR	Virtual Reality/Augmented Reality
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package

1 Executive summary

SHELTER aims at developing a data driven and community-based knowledge framework that will bring together the scientific community and heritage managers with the objective of increasing resilience, reducing vulnerability and promoting better and safer reconstruction in historic areas.

The complexity of SHELTER is due to the great number of interactions between different layers belonging to various scientific, technological, social and human domains at local and global scale. It is impossible to think of tools embedded in a rigid architecture implementing a top-down approach to face all the barriers and constraints of disaster resilience. The solutions must be based on methodologies and standards applied with a certain degree of flexibility that takes into account the local unique characteristics and history, but with a mid/long-term perspectives strategy. Any evolution of data and knowledge systems towards an enhanced digitized public good management, including Cultural and Natural Heritage, should start from existing conditions and available resources in order to implement sustainable and feasible actions for resilience.

The scope of this deliverable is to describe the Task 1.1 (**Identification of data and knowledge sources and integration and interoperability requirements**) activities and outcomes with a focus on the useful data-sources identification, classification and evaluation. To do that, a set of data assessment criteria and data assessment procedure have been defined based on literature review, a list of standards and the inputs from Task 6.1 (**GLOCAL User requirement)** and Task 2.2 (**Systemic resilience assessment and monitoring framework for HA: structure of indicators, definition of KPIs and resilience co-monitoring strategy)**.

The output includes a methodology to collect, filter and analyse the data, a structured list of datasets and knowledge sources identified for specific Cultural Heritage (CH) climate and disaster resilience requirements, the criteria adopted to identify the quality of the datasets and data sources in terms of interoperability and replicability standards compliance and a roadmap to achieve the maximum and durable data exploitation.

The content of this report describes the methodology that has been defined to generate the T1.1 outcome, the description of the data sources classified and evaluated as relevant for SHELTER scope. The link with the D1.2 – **Building of best/next practices observatory** is hereby described as well as the results achieved within T1.2 (**Codification of existing knowledge**). The strategy implemented in T1.2 has been agreed and shared according to the approach implemented in T1.1 and the T1.2 results are considered part of the relevant information and knowledge sources of the WP1. This document also describes the methodologies to share the data, information and knowledge internally and externally the SHELTER consortium, including the replicability, and it ends with a detailed analysis of the identified gaps and a possible suggested roadmap to fill these gaps.

The strategy and the approach proposed in this document aims also at bringing the data beyond SHELTER. The implementation of the data management plan, together with the actions of dissemination and knowledge transfer of data management concepts will reinforce the accessibility, replicability and reusability of the datasets and the knowledge ensuring their operationalization for long after the end of the project.

2 Introduction

The scope of this document is to describe the data and knowledge sources that have been identified within SHELTER for specific CH climate and disaster resilience requirements in compliance with the interoperability standards and to introduce the methodology that has been applied to ensure the data management guidelines compliance. The data survey has been run in the framework of **Task 1.1 – Identification of data and knowledge**, which is integrated in the SHELTER operational knowledge framework as shown in Figure 1 (*red boxes*) and described more in details in D2.1 - HA resilience structure.

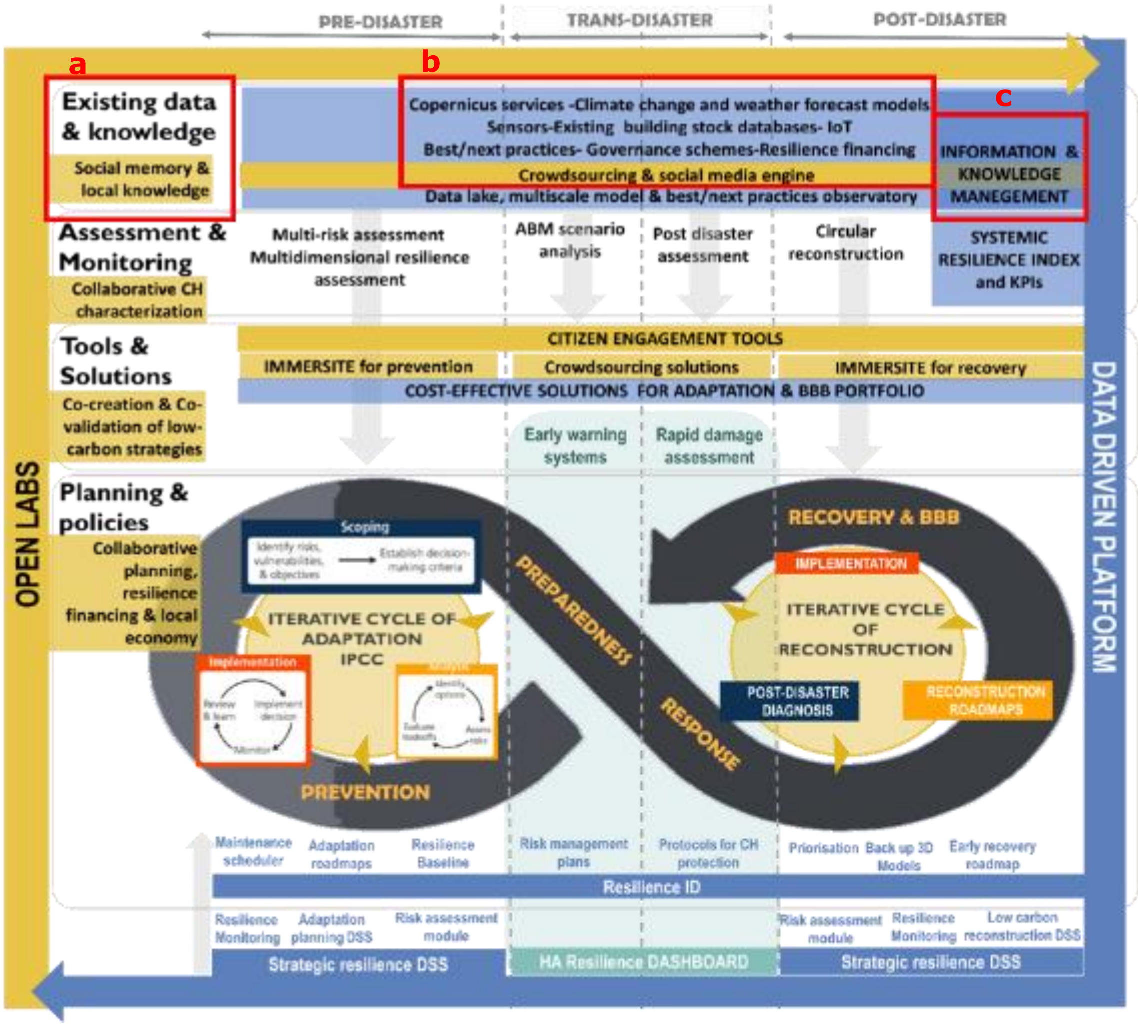


Figure 1 – Task 1.1 contribution to SHELTER Operational knowledge framework (*red boxes*)

The '*Existing data and knowledge*' segment (a) of the SHELTER operational framework (pre-disaster phase) is described in Chapter 3. Here the methodology applied to find the criteria for data selection together with the workflow that generated the Data Mapping

Form (DMF) as a tool for collecting all the relevant information is explained. Chapter 4 describes the Copernicus services (and more in general the selected satellite sources), the climate change and weather forecast models, the Internet of Things (IoT) and the crowdsourcing and social media engine, which can all be transverse with respect to the disaster phases (segment b). Segment (a) also includes the social memory and the local knowledge described more in detail in Chapter 5. Here best and next practices are described as useful knowledge for the trans-disaster phase (segment b).

The *'Information and knowledge management'* segment (c) is developed in the framework of other WP1 tasks (T1.3 and T1.4), but receives as input all the information collected within the T1.1 related segments by structuring all the results through the following methods:

- a Data Lake, for heterogeneous data
- a Multiscale Multisource Data Model, for information retrieved through the case studies
- a Best/next Practices Observatory.

As part of the segment (b) of SHELTER operational framework, a particular relevance is given in this document to the Copernicus Programme and its services (see section 4.2). In fact, as more in details described in the *'Copernicus services in support to Cultural Heritage'* report, 7.5% of the Cultural Heritage user requirements (URs) is already fully covered by Copernicus core services products in their current form, and an additional 19.0% of user requirements is partially covered by existing Copernicus core services products without adaptation. With the support of Sentinels and Contributing Missions capabilities, 50% of the user requirements identified in the report could be fully covered, while an additional 14% could be partially covered. Those partially covered user requirements could potentially be supported by the downstream industry that has access to very high resolution and/or very high revisiting time data.

SHELTER aims at improving the capacity of the Historic Areas (HA) to cope with hazards and Climate Change (CC) related events through the identification and integration of multiple data sources (satellite, sensors, crowdsourcing, predictive models, statistical models…) and existing knowledge (including local social memory regarding past events, best and next practices and results from linked research initiatives).

By a deep understanding of the hazard, the exposure and the vulnerability of the historic area, the local dynamics and the implementation of innovative governance and community-based models, it is possible to provide useful methodologies, tools and strategies to enhance resilience and secure sustainable reconstruction. Due to the information complexity and the diverse data sources, SHELTER framework will be implemented in multiscale and multisource Data Driven Platform (DDP), able to provide the necessary information for planning and adaptive governance. All the developments of the project will be validated in five Open Labs (OLs) across Europe, representative of main climatic and environmental challenges in Europe and different heritage's typologies. These case studies have been selected to maximise the impact of project results, since their combination of diversity and similarities makes them perfect test beds for SHELTER approach and solutions.

SHELTER will generate knowledge regarding the hazards with more impact in CH through these five case studies that cover earthquakes, storms, floods, heat waves, wildfire and subsidence. Three of the cases are multi-hazard situations and two are transboundary. Table 1 shows the five selected Open Labs and the related impacting hazards:

SCALE	OPEN LAB	HAZARD
URBAN	Ravenna - Italy	Earthquakes
		Subsidence
	Dordrecht - Netherlands	Floods/Storms
	Seferihisar - Turkey	Earthquakes
		Heat waves
CROSS-REGIONAL	Baixa-Limia Serra - Spain	Wildfire
	Sava River basin	Floods

Table 1 – SHELTER OLs described in terms of the scale characterizing them and the impacting hazards

Table 2 from D2.1 – **HA resilience structure** shows how catastrophic events may impact on cultural heritage assets, structures and artefacts, with a particular focus on the hazards impacting the five SHELTER OLs.

Type of event	Impact related to cultural heritage
EARTHQUAKES	**Climate change effects:** not related
Physical	One of the natural disasters with the most devastating impact in terms of loss of lives and damage to structures. Frequently, followed by other disasters such as fire, floods, landslides or tsunamis. The case of L'Aquila, Abruzzo (2009, Italy) showed that only 23% of cultural heritage buildings were adequate for earthquakes.
Social	Loss of human lives and abandonment of HA
Environmental	Soil erosion, loss of biodiversity, and, higher risk of landslides
Economic[1]	Decrease in tourism and related activities and economic losses for CH managers and insurance companies.
STORMS	**Climate change effects:** Increase in the number of extreme storm events
Physical	Material decay (fungal growth, degradation of material, biogenic patinas and deterioration of movable heritage). Intense wind-driven rain can alter the distribution of damage on facades. Penetrative moisture into porous CH materials. Static and dynamic loading of historic or archaeological structures.
Social	Loss of identity and common values; progressive abandonment HA due to new comfort parameters. Significant adverse effects on human health, increasing mortality and health risks
Environmental	Adverse effects on vulnerable ecosystems. Decrease in the landscape quality and biodiversity loss.
Economic	Economic losses for heritage managers and/or insurance companies.

[1]Any one of these risks has a huge economic impact in the area, but in this table only the impact directly related with CH is addressed.

FLOODS	**Climate change effects:** Coastal, fluvial and pluvial floods due to global temperature rise (increase of precipitation volume and number of extreme events)
Physical	Damage and failures due to: 1) static and dynamic loads (water pressure, water flow, uplift forces), 2) impacts from floating objects, 3) wetting of building materials, 4) effects of soluble salts, chemical pollutants and biological infection, 5) temperature and humidity fluctuations and hygrothermal cycles increase material decay (cracking, detachment, fungal growth, biogenic patinas…).
Social	Disruption of communities, loss of rituals and breakdown of social interactions.
Environmental	Soil erosion, loss of biodiversity and higher risk of landslides. Deterioration of water quality.
Economic	Decrease in tourism and economic losses due to damages in infrastructure and buildings.
HEAT WAVES	**Climate change effects:** Heatwaves & higher than average temperatures: Prolonged periods of abnormally hot weather and presence of tropical nights (minimum temperature at night of 20 ºC) because of the global warming effect
Physical	Material decay due to temperature and humidity fluctuations, hygrothermal cycles and increasing biological colonisation**.** Deterioration of facades due to thermal stress. Changes in 'fitness for purpose' of some structures (alterations to the historic fabric due to the introduction of engineering solutions)
Social	Progressive abandonment of HA due to temperature increase; impact on health and wellbeing (specially in vulnerable population)
Environmental	Energy demand increase for air conditioning. Increased bushfire risk due to increase in temperature. Heat stress on vegetation, crops and biodiversity.
Economic	Impact on the local economy due to depopulation, lowered labour productivity and changing tourism patterns.
WILDFIRE	**Climate change effects:** Increase in global temperatures, heat waves and related to water scarcity/droughts
Physical	In European countries, fire is the common catastrophe that threats built heritage in urban areas and fire prone cultural landscapes. Damage due to a water-based suppression system. Air quality which may potentially impact CH.
Social	Depopulation. Increase health and mortality risks due to health hazards from air pollution.
Environmental	Forest fires are a central part in shaping forest ecosystems but can be particularly devastating especially when they threaten residential areas. Air pollution.
Economic	Economic losses for heritage managers and/or insurance companies. Indirect economic losses can also occur through job losses and lower incomes, especially in areas that rely on offering tourist products that are based on natural and cultural beauty.
SUBSIDENCE	**Climate change effects:** triggered by various extreme weather events.
Physical	Damage in residential, commercial, and public buildings, spaces and assets and public utility networks
Social	Fatalities and population displacement
Environmental	Removal of vegetation and soil and the deposition of this material on different ecosystems
Economic	Damage to economic assets and transport infrastructure and have a negative effect on the economy both directly and indirectly (from job losses and restricted ability to access work).

Table 2 – Impact of CH events (Source: D2.1 – HA resilience structure –*updated version*)

To test the community based replicability and the data driven approach, the case studies have been selected considering:

i) their CH value and diversity (from archaeological sites to cultural landscapes including urban and rural, cultural and cultural/natural heritage),

ii) their especial exposure to diverse hazards,
iii) their geographical representability and climate conditions,
iv) their scale and typology (from building to transnational regions),
v) their full commitment with SHELTER,
vi) their diverse governance schemes and participatory experiences,
vii) their different level of information infrastructures

With respect to the first criteria of selection, the Baixia-Limia Serra and the Ravenna OLs represent a valid example of two different types of Heritage, Cultural and Natural respectively (see the '*Copernicus services in support to Cultural Heritage*' report):

- **Cultural Heritage (CH)**: including Tangible and Intangible Heritage (TH and IH). TH comprises buildings and historic places, monuments, artefacts, etc., which are considered worthy of preservation for the future. These include objects significant to the archeology, architecture, science or technology of a specific culture. Moreover, TH includes places with significant heritage value, because of its special association with the life or works of a person (or group of persons), for its importance or events in cultural history or for its association with people, events, places and themes. Anyway, cultural heritage does not end at monuments and collections of objects. It also includes traditions or living expressions inherited from our ancestors and passed on to our descendants, such as oral traditions, performing arts, social practices, rituals, festive events, knowledge and practices concerning nature and the universe or the knowledge and skills to produce traditional crafts (IH).

- **Natural Heritage (NH)**: natural features consisting of physical and biological formations or groups of such formations, which are of outstanding universal value from the aesthetic or scientific point of view.

As a consequence of this difference, also the kind of information and data required can be different, depending on the user communities involved:

- the Natural Sciences user community, in the case of Natural Heritage;
- the Cultural Heritage professional user community, in the case of Tangible and Intangible Heritage.

Together with these communities the National, Regional or Local authorities, the site operators, the urban planners as well as the intermediate user communities may intervene in one of the segments around which the Cultural Heritage is structured, but are usually transverse.

The CH user communities have different demands for performing their activities, which are expressed in the form of user needs. Being SHELTER especially focused on monitoring the Cultural and Natural Heritage, the following user needs are in particular of interest for the project:

- Monitoring the evolution of the natural environment of both Cultural and Natural Heritage
- Observation of damage on the built structure of a CH site
- Drawing of conclusions to facilitate an emergency intervention

There are many similarities between the user needs related to CH and NH, as they share common objectives and features. However there is a gap at this level between how NH is monitored compared to CH. Indeed, almost all European NH sites are covered and follow specific processes of preservation, while for CH, the definition and implementation of conservation management plans has not always been generalised in practice. As a general rule, management plans for CH need to be necessarily generalised at first, taking adavantage from the already existing conservation and preservation guidelines. Then management plans will be tailored according to the different government budgets, cultural approach, threats, etc., that characterize the different countries object of conservation.

In this respect, both the Sava River basin and the Baixa-Limia Serra have shown a quite advanced level of available information in terms of Disaster Risk Management and hazard monitoring already before SHELTER was launched (see SavaGIS platform for example), at least for what concerns the preservation of the related natural areas. Indeed, an effort to integrate such information with attributes data related to the CH component has been necessary. Under request of the International Sava River basin Commission (ISRBC), as representative of the Sava local stakeholders, a mock-up for defining a CH attributes template has been designed in the framework of Task 1.2 (see D1.2 - Building of best/next practices observatory, section 6.7.1 and section 4.6 of the present document). Such a mock-up has the scope to provide a format for collecting and storing data, information and knowledge that will be extracted from relevant institutions.

Providing site operators with systematic accessible data is key to facilitating the definition and implementation of site management plans. In fact, even if data is available for European Natural Heritage sites, it appears that site operators are working on Cultural Heritage lack data and thus they cannot implement or share good practices.

SHELTER user communities need to map the cultural landscape of the site and identify the specific risks it is exposed to, and observe the damage on the built structure of a CH site. This requires updated information on land use of the surrounding area, the evaluation of the site’s exposure to all potential risks (*e.g.* geo-hazards) because of its location, positioning, and surroundings, the analysis of the material composition of visible parts to understand the overall structure and identify potential damage. The above mentioned mock-up has been designed so as to reflect these user communities needs.

By comparing needs in both NH and CH environments and identifying their similarities, one key conclusion is the coherence between this more integrated approach in which CH could benefit from best practices from the NH community, in order to foster its global development and sustainability.

In a nutshell, through the operationalisation of the existing heterogeneous data and knowledge, Task 1.1 serves SHELTER in reaching what is considered its ultimate scope: the implementation of a Sustainable Historic Environments hoListic reconstruction through Technological Enhancement & community based Resilience.

2.1 Relation with the other activities in the project

As the other SHELTER Work Packages (WPs), WP1 is strongly related directly and indirectly to all the other activities of the project. In particular, being the scope of WP1 **the operationalization of existing data and knowledge**, most of the actions within this WP are focused on the identification and the transformation of the existing, spare, unstructured, unstandardized, unfiltered and heterogeneous data and information into data and knowledge fully accessible, usable and exploitable both within and outside SHELTER (see Chapter 3, 6 and 7).

The WP1 is built around the concept of data and knowledge sources identification and exploitation. The implementation of these two pillars represented by '**identification**' and '**exploitation**' can be carried out by mobilizing expertise and knowledge resources inside and outside the SHELTER consortium. The interactions with the other WPs are described in Figure 2:

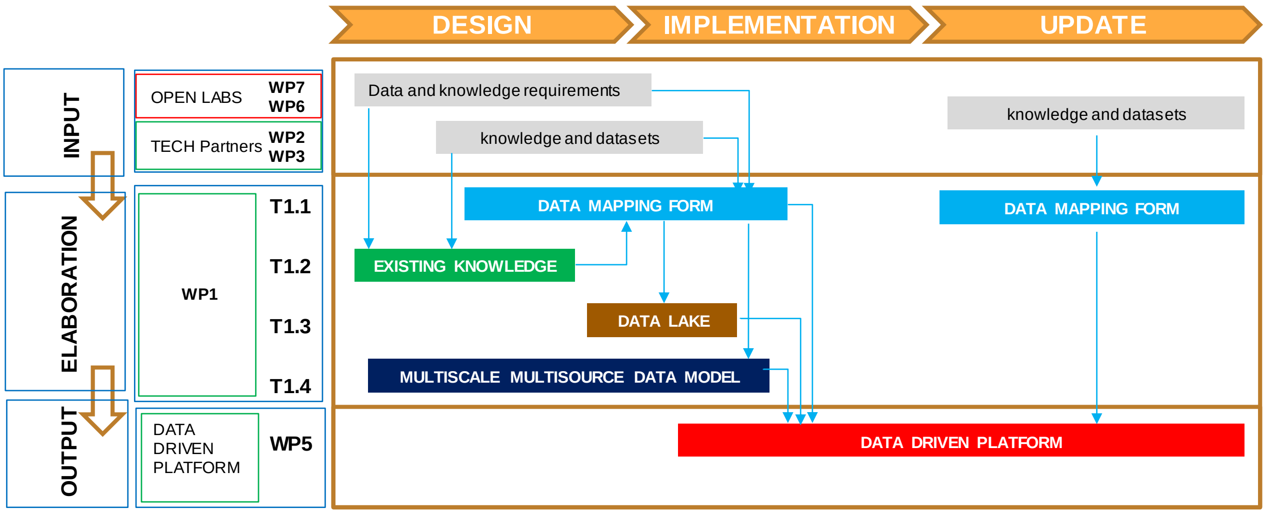


Figure 2 – Interactions between WP1 and the other WPs in SHELTER

Data and Knowledge to be exploited and operationalized in SHELTER can be identified from three main sources:

- The **Open Labs**, which represent the so-called '*problem owners*'. They are supposed to have all the local and detailed data and knowledge that must be collected and shared within SHELTER. The OLs also expose the specific requirements and expectations which are indispensable to filter properly all the data collected from general data sources and to return them customized, useful and usable technological solutions. In Figure 2 the OLs inputs are indicated as from **WP7** that coordinates all the OLs activities and as from **WP6** that is devoted to the OLs requirements collections.
- The **SHELTER technological partners**, which represent the solution providers. Within SHELTER, the technological partners generate both tools and outcomes in form of maps, data, indicators, methodologies to solve the problems identified by the OLs. This kind of data source is identified as expected input from **WP2** Knowledge generation: Systemic HA resilience assessment and monitoring, and **WP3** Tools and solutions for prevention, preparedness, response and recovery.
- The **external data sources**, which represent a relevant complementary added value to the technological solutions provided in response to the problems owned by the OLs. The external data sources category includes for example the Copernicus data and services like Earth Observation data, advanced products, climate models output, climate scenarios and regional seasonal climate forecasts. The external data sources are considered **WP1 internal input** since they are identified thanks to the expertise of WP1 partners.

The WP1 internal relationships between the four tasks are described in the Figure 3:

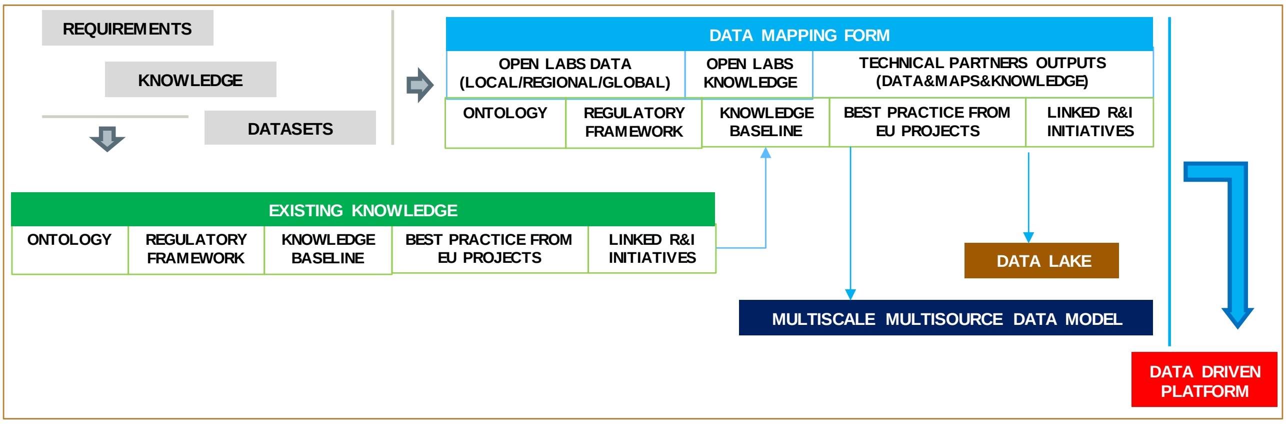


Figure 3 - WP1 internal relationships between the four tasks of which is composed

The data and knowledge requirements are taken as input of both the tasks devoted to data and knowledge identification, which are the **T1.1** for what concerns the data and **T1.2** for what concerns the existing knowledge. Within these two tasks, datasets and existing knowledge belonging to the three categories of sources above described are identified, explained, filtered and assessed with respect to the requirements, the indicators and the expertise of WP1 partners. The datasets and the existing knowledge are described by means of structured data description tables that will converge in the Data Mapping Form (see section 3.1). Both the structure and the datasets are considered as input for the Data Lake design (**T1.3**) and the Multiscale Multisource Data Model design (**T1.4**). With this approach the data structures are tailored to the variety and the characteristics of the data and the knowledge that will be operationalized within the project.

The output of WP1 is addressed to the whole project but, in particular, it represents the main input for the **Data Driven Platform** of WP5. Indeed, datasets, existing knowledge, Data Lake and Multiscale Multisource Data Model are all fundamental pillars upon which the Data Driven Platform must be designed, developed and implemented.

2.2 Document structure

This document is comprised of the following chapters:

Chapter 1 contains the project and the document overview.

Chapter 2 presents an introduction to the document.

Chapter 3 presents the methodology applied for data and knowledge collection.

Chapter 4 describes in detail the different kinds of data sources (satellites, Internet of Things, crowdsourcing and social media, socio-economic data sources and local data sources) and the criteria followed to select them.

Chapter 5 is dedicated to the description of the knowledge considered useful for CH climate and disaster resilience (past events, best and next practices, linked research initiatives)

Chapter 6 contains information on how the selected data are in compliance with the replicability standards (metadata standards, ontology).

Chapter 7 identifies the gaps in the data and knowledge gathering and defines a feasible roadmap, also describing knowledge that has been shared among partners for SHELTER internal use.

Chapter 8, here the main conclusions are drawn.

Chapter 9 provides the references.

Chapter 10 contains the annexes (questionnaire for data info collection, the Data Mapping Form template with the related instructions and naming convention, the DMF evolution document and the survey for collection on IoT information.

2.3 Contribution of partners

Table 3 details the contribution of each partner:

Partner	Contribution
SISTEMA	Responsible for the coordination of the task and deliverable. Responsible for definition of the overall approach and methodology. Drafting of all sections
EGIS	Contributor to Section 3.2: Data Models
TECNALIA	Contributor to Section 4.2: The Copernicus programme and services
RED	Contributor to Section 4.3: IoT data
LINKS	Contributor to Section 4.4: Crowdsourcing and Social Media data
UMAS	Responsible for contents of Section 4.5: Socio-economic data sources

Table 3 – Contribution of partners

2.4 Applicable and reference documents

The applicable documents are listed in the table below:

ID	Document	WP
AD.1	D1.2 - Building of best/next practices observatory	WP1
AD.2	Wiki Page	WP1
AD.3	D2.1 - HA resilience structure	WP2
AD.4	D6.1 - GLOCAL_V1.0	WP6
AD.5	D6.5 – Methodology for Local Knowledge Extraction	WP6
AD.6	D9.3 - DataManagementPlan_V1	WP9

3 Methodology for data and knowledge collection

The overall objective of SHELTER is to establish a cross-scale, multidimensional, data driven and community based **operational knowledge framework** (Figure 1) for heritage-led and conservation-friendly resilience enhancement and sustainable reconstruction of historic areas to cope with climate change and natural hazards.

This operational knowledge framework will be the result of the interplay of two processes collaborating to feature individual solutions for each HA:

i) Open Labs approach that provides a continuous framework for local knowledge extraction, citizen´s engagement, co-creation, capacity building and innovation
ii) a Data Driven Platform that supports diagnosis, decision making, implementation and monitoring based on existing knowledge and heterogenous data

In order to reflect such approach while performing the knowledge operationalization in Task 1.1, several steps have been necessary: firstly, existing data, information and (local) knowledge have been identified; secondly, the non-existing knowledge for assessment and monitoring that will be generated during the project and the required tools and technologies that will be adapted or generated have been detected.

The logic workflow for this action is strongly connected to the list of Key Performance Indicators (KPIs) foreseen in Task 2.2 **- Systemic resilience assessment and monitoring framework for HA** (structure of indicators, definition of KPIs and resilience co-monitoring strategy). Such list represents the source for a harmonise and multiscale indicator based risk dependent resilience assessment built on hazard, exposure and vulnerability (of single risk and combinations of risks) at artefact, building and district scale, but also generalised multiscale resilience (from building to transnational regions). The workflow for data and knowledge collection has been adapted to run in parallel with the T2.2 flow of information and it started from the analysis of data and knowledge already available for the Open Labs that are objects of SHELTER with relation to the identified hazard (Table 1).

To collect this information in a structured way, a matrix acting as a canvas for the project developments first, and as guideline for replication in other cities posteriorly, has been further designed, namely the **Data Mapping Form** (DMF). A dataset description template was preliminarily outlined in D9.3 – DataManagementPlan_V1 (see Annex A of D9.3 and the related Sharepoint link: Data Mapping Form), but through the creation of the DMF, the Data Management Plan has been put into practice, complying with the main principles therein described.

The DMF is meant not only for geospatial data, but also for information and knowledge not directly referring to a position in space and time. Among this information, it is particularly crucial to carefully structure the data related to Intangible Heritage (see

Introduction). In fact, due to its intrinsic nature, this kind of Cultural Heritage is very difficult to gather and to describe, because it includes a wide range of heterogeneous information, with various grades of impacts and reliability (see D6.5 – **Methodology for Local Knowledge Extraction**). Anyway, thanks to the scalability characterising the DMF, it will be possible to also include this kind of knowledge. To this scope, a searchable metadata structure should be implemented: the aim in this case would not be to describe the IH itself, but to register its existence through the use of dedicated tags, for example. Thanks to the way how the DMF is designed, it will be then possible to query the developed database and to perform many different types of analyses (*e.g.* to calculate the percentage of IH per OL).

In general, the different sections that compose the DMF partly reflect the user requirements identified in the '*Copernicus services in support to Cultural Heritage*' report, which have of course been tailored on SHELTER specific needs. It represents in fact the translation of the user needs considered as relevant for SHELTER into user requirements which can be characterised by the provision of information on (see Annex I):

- **Data description** (data type, status of the data – existing or foreseen)
- **Data collection** (format, size, time and area coverage, spatial and temporal resolution, update frequency, collection information, licence, ownership and author)
- **Data access** (access mode, restrictions, links and metadata)
- **Data storage and processing**
- **Data analysis**
- **Data tools** (for processing and/or analysis)
- **Value creation** (data application field, hazard type and data end-users)
- **Future data**

This information represents the main sections of the DMF, with each section being related to a specific group of steps in the data value chain. Each of the fields that compose the DMF has the objective of creating standardised metadata thus allowing the application of the FAIR principles (Findable, Accessible, Interoperable and Re-usable) (Wilkinson, M. *et al*. (2016) [1] and the INSPIRE standards - see D9.3: DataManagementPlan_V1). Table 4 shows the match between the FAIR principles and the main DMF fields:

FAIR principle	**DMF fields**
Findable	description, type, format, collection, license, ownership/author
Accessible	access mode, access restrictions, access links, metadata
Interoperable	format, metadata standard
Re-usable	license

Table 4 – Match between the FAIR principles and the main DMF fields

This template is used not only for recording data information and descriptions, but also as a SHELTER internal tool to support the communication and the info exchange among partners, particularly between technology partners and OLs or among technology partners. With respect to the Data Driven Platform development, it is necessary for the technology partners to be continuously updated on the type of data available from the OLs, in order to properly meet the technical requirements of each tool composing the DDP. In addition to the already existing data and knowledge in fact, the DMF is used to collect information about data foreseen to be produced by technical partners involved in SHELTER according to the tasks described in the Grant Agreement. The detailed description of such data and knowledge in a structured format is also fundamental for WP5 tasks, in order to design, implement and integrate proper tools and solutions. This is for example the scope of the '*IoT*' field in the DMF (indicating whether a certain dataset comes from an IoT source or not) or of the '*Data example*' field (containing a link to each data example available for WP5 partners convenience).

Thus, while the collection of information was initially triggered by WP1 leader in the role of intermediate between the OLs and the technical partners, as soon as the DMF has taken an advanced form and all the SHELTER partners have been trained on how to use it, it has started being utilized as internal tool also for other tasks scope, beyond the Task 1.1. Through the several informative fields which the DMF is composed of, the possibilities of performing dedicated queries are numerous. It is sufficient to apply the correct filter, depending on the specific need.

The DMF represents a living document, being regularly reviewed and updated in its content and structure during the course of the project. This continuous evolution also depends on the fact that the parameters taken into account for identifying and preventing potential risks (*e.g.* climate-related risk) change continuously and the user communities need to conduct a constant environmental (climate change, geo-hazards, pandemic events, etc.) monitoring.

In order to keep track of the changes involving the DMF, a supplementary document (DataMappingForm_Evolution.doc, Annex II) has been drafted to collect future updates on the basis of the considerations derived from reviewing the already described dataset (see section below).

Both the DMF and the related documentation are available on Sharepoint for internal use of the SHELTER consortium. The DMF is also accessible through an external link.

3.1 The Data Mapping Form

The workflow, the principles and the methodology applied to design and further develop the DMF are hereby described and schematically shown in Figure 4:

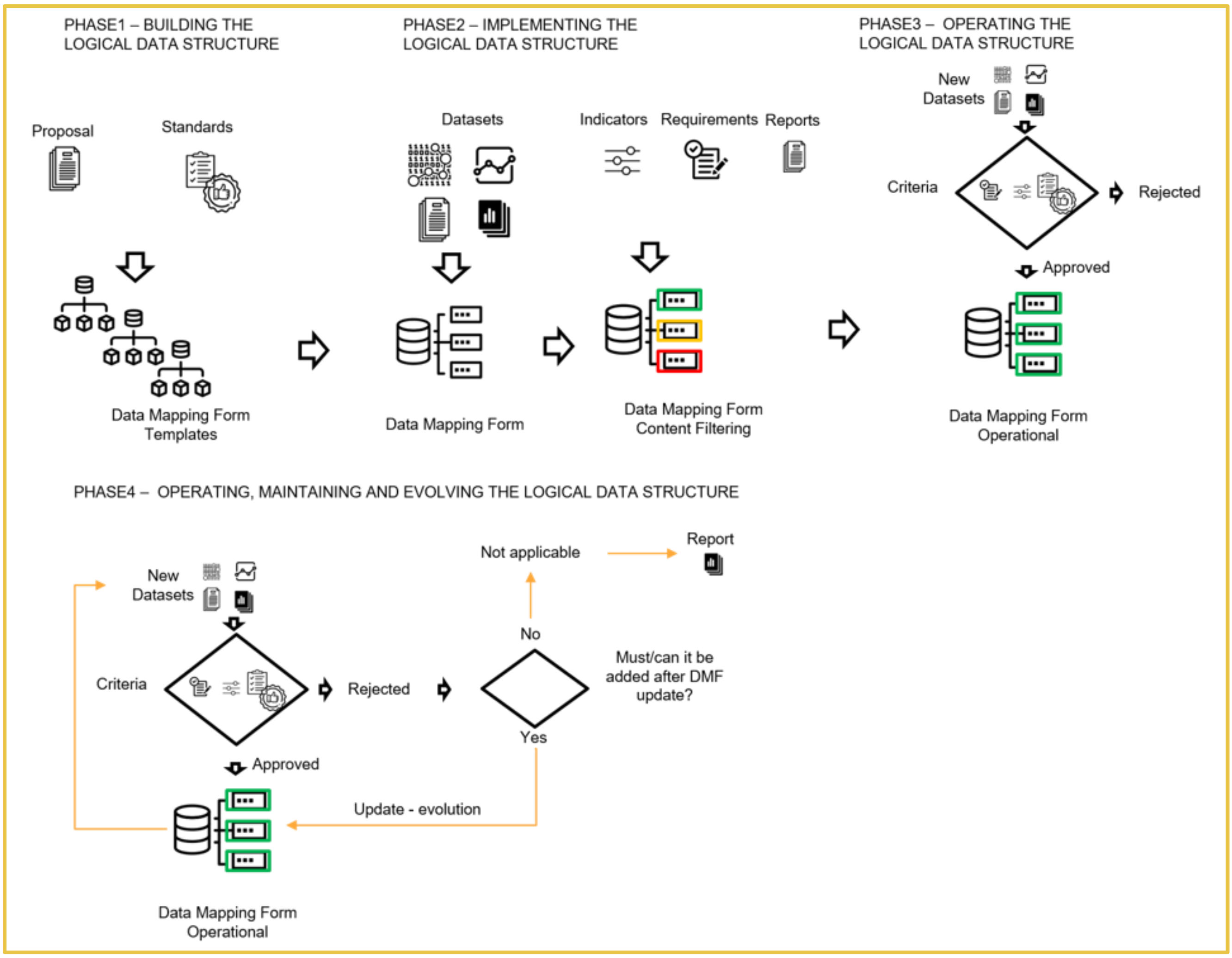


Figure 4 - Task1.1 methodology phases

- **Phase 1 – building the logical data structure:**

The list of standards and the inputs expected from T6.1 **GLOCAL – User requirement** was originally considered as the basis to start collecting stakeholders' requirements and a dedicated questionnaire was drafted as complementary to the one prepared by WP6 partners (see Annex III). Such a questionnaire was meant to collect preliminary information about data types and format, data accessibility, ethics and legal compliance, and data application field. After circulating the questionnaire among the partners and having collected the preliminary feedback from them, a bilateral interaction between WP1 and WP5 leaders has started, in order to build the logical data structure to be applied for collecting the needed information.

During this phase the template of the Data Mapping Form was drafted in compliance with the FAIR principles (Findable, Accessible, Interoperable and Re-usable) and the INSPIRE metadata standards (see D9.3 – DataManagementPlan_V1). Thus, the main sections previously described within this document were defined on the basis of the user requirements identified through the project proposal. Each section was then expanded in a matrix, according to the guidelines received from WP5 partners (see Annex I, Figure A- 1), particularly for what concerns the INSPIRE metadata standards, which were inserted as external sheet to list the minimum set of metadata elements necessary to comply with Directive 2007/2/EC (see Annex I, Figure A- 2).

At the end of this development phase, the matrix looked as composed by 7 sheets:

- the **INSPIRE metadata** standards sheet
- the **Instruction** sheet, to facilitate the comprehension of the different fields composing the DMF and to guide each user during the filling process (see Annex I, Figure A- 3). The instructions are characterised by different colours to highlight the position in the data value chain and come with examples or with a list of allowed values where necessary. This sheet was updated with new instructions each time a new field was introduced in the DMF during the design process. It contains instructions tailored both to the OLs and the technology partners requirements. Through the continuous interactions with them in fact, it was possible to refine the DMF according to the different needs they expressed. All partners were guided through the DMF compilation by applying a very flexible approach, which is that expected contributions don't always have to perfectly match all the DMF fields. Being structured for collecting both data and knowledge information, most of the DMF fields in fact are supposed to contain an indicative value. This is the case for example of the '*Time coverage*' and '*Area coverage*' fields, which can result ambiguous when it comes to describe a climate report, but that can be interpreted as the area to which the report is applied and the validity time of the report respectively.
- **Five Open Labs** sheets, to collect the needed information on separate tables, one per OL, being the five OLs different in terms of hazards impact, geographic scales and all the aspects already described in the Introduction of the present document.

- **Phase 2 – implementing the logical data structure:**

  During this phase the DMF was filled with preliminary data information in order to check its robustness. A naming convention was established (ID) based on Task and Subtask code related to the described dataset (*WP#Task#Subtask#Progressive#*). Once the tool was internally validated, all SHELTER partners were asked to contribute with descriptions of further datasets considered useful for the scope of the project, thus allowing another round of validation. During this process many interactions occurred, especially with OLs referent partners and several inputs were collected in order to make the DMF a more powerful tool for SHELTER scope.

Thus, new fundamental fields were added, such as:

- **IoT**: to indicate whether the described dataset derives from IoT sources. This way the technology partners in charge of developing the IoT service for real-time inside the DDP, for example can have a clear overview on which kind of data are expected to feed such service
- **Hazard type**: to match the described dataset with the type of hazard it can be applied to. One dataset can be used in fact to describe or to prevent one specific hazard (*e.g.* a seismic risk map) or can be related to more than one hazard type (Multihazard dataset, such as the *EM-DAT – The International Disaster Database*. Other datasets can be non-specific (*e.g.* topography maps), but they play an important role when used in conjunction with hazard-specific datasets
- **DRM phase**: to indicate which of the four disaster risk management phases the described dataset is applicable to. This field is extremely important for Task 1.1 scope as the operational knowledge framework is the result of the intersection between the four Disaster Risk Management (DRM) phases (prevention, preparedness, response and recovery including reconstruction) and the tools and mechanisms that support the resilience building in HA (existing data and knowledge operationalisation, assessment and monitoring framework, tools and solutions development and collaborative planning)
- **Data example**: at least one file example per each described dataset was collected in order to provide WP5 partners with the basic information in terms of data format, data size, data accessibility, and so on (see Annex I, Figure A- 2 for the example file naming convention). This field is particularly important for what concerns the microservices that are going to be developed within the DDP. As previously explained in fact, together with data available from the OLs, the DMF is used also to collect information about datasets that are foreseen to be generated by the technology partners within SHELTER. To this scope further sheets were added, one per technology partner (named as *P_partnerID_partnerShortName*) and each technology partner was asked to fill in his/her own sheet with description of the dataset linked to the related SHELTER tasks. A preliminary interaction with the involved partners was triggered by extracting from the proposal the relevant actions foreseen for each Task and Subtask. The descriptions of such actions functioned as a reminder for the technological partners, who were then expected to convert the respective actions into

production of datasets or development of tools. This conversion, for many of the technological partners, is still in progress.
- **Task/Subtask, Input task, Input direction**: these fields were added only for the abovementioned technology partner sheets to indicate the ID of the related task, the ID of the Task that will provide input for producing the described dataset and the input direction respectively. This should facilitate the WP5 partners in identifying the correct inputs and outputs with respect to all tools and solutions to be developed in SHELTER.

A preliminary review of the DMF was performed once:

- a minimum number of records were inserted for each Open Lab sheet
- at least the sheets dedicated to existing/foreseen data produced by technical partners directly involved in the DDP development were filled with the necessary information

The review was performed through the support of WP5 partners, so that the DMF could reach a more consolidated version.

Section 3.1.1 describes in detail the criteria (both technological/functional and content-related) that have been applied.

- **Phase 3 – operating the logical data structure:** with this phase the DMF has become operational and all new datasets possibly further described have been subject to the above mentioned criteria of selection. Indeed, this phase implies a repeated and continuous data review process, with particular focus on the list of indicators developed in Task 2.2, versus the user requirements defined in Task 6.1 (see section 3.1.1.2 and 4.1) and the dataset described in the DMF. In particular, the list of indicators classified within T2.2 on the basis of a RACER evaluation (see section 4.1) has been used for identifying the datasets relevant for the project. The indicators with total RACER score = 3 (*i.e.* the highest score) have been initially selected to investigate on available linked data. In the next steps also the rest of the indicators will be taken into account and a further review of available data will be performed once the indicators will be validated by the OLs.
  Those indicators and user requirements that couldn't find any match with the available data are going to be included in the data gap analysis (see section 7).

- **Phase 4 – operating, maintaining and evolving the logical data structure:** datasets approved after the selection process will represent a valid output for components developed in WP5 (Data Lake, Data importer and mapper, Resilience dashboard). Starting from this phase, the DMF can be used in this sense to identify the matching between the selected data and the tools that are going to be developed in SHELTER (WP5). To this scope a new sheet was introduced in the

DMF by WP5 partners, namely the '**Tool-Matrix**' sheet (see Annex I, Figure A- 4), which is used internally by WP5 contributors for associating the SHELTER Tools&Solutions to the DRM phases and to the OLs in which they could be demonstrated. The link between this matrix and the DMF is represented by the field '*data IDs*', containing the ID of each dataset described in the DMF and representing an output of the related tool/solution.

The described methodology for the DMF design and implementation is based on continuous interactions with SHELTER partners. In order to keep track of the several changes made to the DMF during the course of the project and of the foreseen integrations, a document was included as annex to the DMF and shared with the consortium on SHELTER repository: the DataMappingForm_Evolution (see Annex II). This document was drafted to collect future DMF updates on the basis of the considerations derived from reviewing the already described datasets. In fact, the DMF is a living document and for this reason it can be subject to further changes/improvements during the course of the project.

Any involved partners are allowed to give contributions, by indicating name of contributors and date. The WP1 leader will evaluate the proposed change in coordination with WP5 partners and will integrate the new suggested update in case of positive feedback.

One of the following status definition has to be used to mark the suggested updates:

a. To be evaluated
b. To be integrated
c. Added

The following section illustrates the different criteria applied for the data selection collected through the DMF.

3.1.1 Data filtering criteria

3.1.1.1 Technological and functional criteria of selection

Figure 5 shows the main technological/functional selection criteria applied to the datasets with respect to the WP5 requirements:

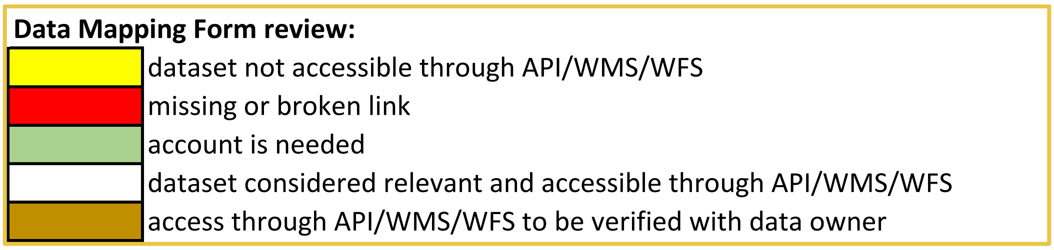


Figure 5 - Criteria of selection applied for the DMF review during the methodology phase 2

All datasets accessible through services that expose **Application Programming Interfaces (APIs)** are selected as considered useful for the DDP implementation. The data retrieval through APIs has in fact many advantages, including the possibility of fetching web data directly, with no other intermediaries needed. Moreover API services allow on-demand data access and are easy to integrate. Also data sources making data available through **Web Map Services (WMS)** or **Web Feature Services (WFS)** are prioritized, because of the many advantages that such services bring both to the users and to the SHELTER technology partners (lead times reduction, cross-collaboration improvement, low cost barrier to entry, connection of multiple data sources, support of real-time spatial analysis, etc.). Of course there are differences among the five Open Labs for what concerns the availability of data characterized by WMS or API and the accessibility of the respective data. Figure 6 shows how the data described so far in the DMF for the five Open Labs are distributed in terms of the abovementioned technological criteria of selection:

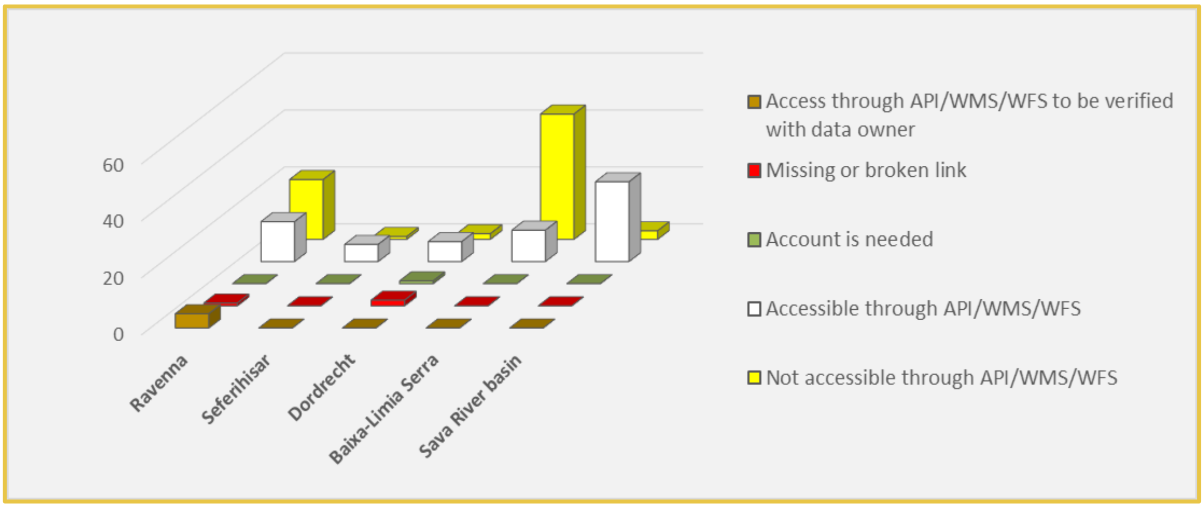


Figure 6 – Distribution of the selected dataset in terms of technological criteria for the five SHELTER OLs

In the above figure both the Ravenna and the Baixa-Limia Serra OLs seem characterized by a quite high number of datasets not accessible through API, WMS or WFS. This is mainly due to the fact that for these two Open Labs there is a lot of unstructured data (such as papers, reports, project descriptions) which cannot of course be accessible through the above mentioned services, but have to be taken into consideration as well, because of the crucial information they embed. Where possible, the raw data which are at the basis of such unstructured data will be searched and the respective data source link will be reported in the DMF in place of the document derived from them.

From Figure 6 it is also clear how the availability of relevant datasets looks pretty scarce for Seferihisar OL. As better explained in section 4.6, this is due to some issues encountered while interacting with the local stakeholders. Anyway, looking at the global distribution of the dataset accessible through API, WMS or WFS (which are the type of data of main interest for SHELTER technological purposes), the data availability seems to be well balanced among the five OLs, with Sava River basin OL being at an advanced status in this sense. This represents of course just one of the interpretations that can be given looking at the overall data distribution, also considering that the situation described by Figure 6 has to be considered as a snapshot taken at the moment of drafting the present document.

Both **open datasets** and datasets with restricted access are taken into consideration, but the former are prioritized because of their capacity of meeting the principles of Interoperability, Re-usability, Replicability and Scalability.

As already addressed in D9.3 – DataManagementPlan_V1 the large amount of data needed in SHELTER cannot be collected into a single repository, because of the almost unlimited needs both in terms of ICT resources and effort typically required by the Big Data challenge of maintaining, operating and updating the datasets. For this reason, preference is given to **linking data** rather than collecting data. This approach will be applied also to those data that are going to be generated within SHELTER. This is connected to the aggregation of scientific knowledge that is more and more performed through data aggregation services and platforms which allow the access to collected raw data. For this reason, the datasets described in the DMF should always come with the related source link (either a direct link or a link to the data provider). Thus, during the data filtering process, each dataset link is checked, and missing or broken links are notified to the responsible partners in order to fix them or discard them where necessary. When a group of datasets is described under a unique data provider (*e.g.* Copernicus) the related link is moved to the field '*Comments*' and each dataset from the group is listed and described separately, with the link for accessing the single dataset indicated as '*Access link*'.

The datasets with **high spatial resolution and local scale** are prioritized. This is linked to the regional character of the selected case study areas. The introduction of Big Data in Disaster Risk Management dramatically reduces the time needed to collect data, to increase the spatial resolution of maps, and to target data more precisely at groups and questions of interest. Anyway, it is still challenging to retrieve the needed data at such high spatial resolution as the one needed for some SHELTER use cases. As a consequence satellite data are taken in consideration together with data from local sources.

Another criterion of selection is based on the availability of **processed versus unprocessed data**. For both Natural and Cultural Heritage, non-technical users, such as international organisations that intervene on the monitoring of sites, need to access processed data that is sufficiently comprehensive and informative for them, and for which IT skills are not required. In the particular case of Natural Heritage, needs are for both current information on biodiversity and vegetation density that provide a very large spatial coverage and for data that provides an understanding of climate change adaptation. Environmental data such as temperature changes and levels of humidity in the atmosphere for example are thus particularly key (see the '*Copernicus services in support to Cultural heritage*' report).

3.1.1.2 Content-related criteria of selection

A fundamental criterion of selection in terms of data content is based on the list of indicators defined in the framework of Task 2.2 and of user requirements identified in Task 6.1 (see section 4.1). For what concerns the **indicators** considered relevant for SHELTER scope, the selection started by firstly taking into consideration those indicators to which the higher RACER evaluation was assigned (i.d. *3 = meets the criteria totally*) by the SHELTER partners in charge of evaluating them (Table 5 and Table 6 of section 4.1 for environmental and statistical indicators respectively). Where possible, a dataset representing an indicator itself is prioritized, such as the *Daily air quality analyses and forecasts for Europe* from the Copernicus Atmosphere Monitoring Service (CAMS) (dataset *72047* of the DMF, see section 4.2.2). If no match is found between the referent indicator and the available dataset, then the data representing the parameters useful to calculate such indicator is selected. This is the case for example of the *24-hour accumulations of precipitation over the Netherlands* for Dordrecht OL (dataset *74010* of the DMF), which can be utilized to calculate the *Daily maximum precipitation corresponding to the return period T* indicator, or the *ERA5 daily maximum temperature of air at 2m above the surface*, which is the basis for retrieving most of the heat waves characterisation indicators. This reflects the concept embedded in the intrinsic definition of '*indicator*', which is the combination of different main environmental parameters, in the case of hazard characterisation and environmental resilience, and of CC and CH data in the case of cultural, social, governance and institutional and economic resilience. In general, reanalysis data (*e.g.,* UERRA, ERA5…) from the Copernicus Climate Change Service (C3S) are particularly useful in terms of content of variables (see section 4.2.2), as each product makes many different variables/indicators accessible at once.

With regards to the **user requirements** identified in Task 6.1, the list of answers collected through the questionnaires that was submitted to the stakeholders during the GLOCAL workshop has been reviewed, particularly for what concerns the DRM questions about data and information (see Figure 1 of this document and D6.1_GLOCAL_V1.0, Figures 14 and 16). As a result of these questionnaires, data and information already available were identified by Task 6.1 partners. Such available data and information were further classified in terms of usable/non-usable for the DMF filtering procedure. A dataset marked as usable is characterised through the above mentioned technological/functional criteria of selection, particularly the accessibility through API services, the open data and the active access link.

QUESTIONS	OL REGION	BEFORE EVENT	AV	DURING EVENT	AV	AFTER EVENT	AV
Which **informations** do you need for an effective risk management? Please name also informations you don't have available!	RAVENNA	monitoring system collecting all the information about the site		monitoring system collecting all the information about the site		monitoring system collecting all the information about the site	
		monitoring system collecting all the information about the risks (frequency, probability, magnitude, etc)		monitoring system collecting all the information about the risks (frequency, probability, magnitude, etc)		monitoring system collecting all the information about the risks (frequency, probability, magnitude, etc)	
		stakeholder database with the related competences		stakeholder database with the related competences		stakeholder database with the related competences	
	SAVA RIVER BASIN	Available observed and forecasted hydrological and meteorological data	1	Available observed and forecasted hydrological and meteorological data	1		
		Available data and information on CH sites and structures	1	Dissemination of warning messages describing what is happening, forecast and the expected impact, as well as advise what action should be taken or trigger a particular emergency response in the emergency plan.			
		Available warnings on vulnerable CH sites and structures					
		Available flood hazard and risk maps of vulnerable CH					
		Available flood risk management plan	1				
		Available emergency, evacuation and communications plan					
	DORDRECHT	Location of Cultural heritage objects	1	contact information of owners cultural heritage		total amounts of damages	
		Vulnerability of Cultural heritage				exact causes of damage	
		Flood data (return period, height, velocity, water quality	1	Flood data (return period, height, velocity, water quality	1	measures taken before flood event	
		contact information of owners cultural heritage				type of cultural heritage	
		long term models of flood damage	1				
		suitable adaptation measures					
	SEFEREKO	probability of risk - not available for all cases		feedback of victims to the hazard		detailed information of damage	
		action plan with spatial instructions - not available to a good quality		action plan with spatial instructions - not available to a good quality		action plan	
	GALICIA	fire prevention plan	1	Roads open for evacuation	1	landowners	
		areas of high risk	1	Roads open for emergency service	1	which species should be used to reforestation	1
		fire prevention actions	1	population living in the area affected	1	is the reforestation working? (but we don't have anyindicators)	1

Figure 7 – Answers to Task 6.1 DRM question about information (D6.1_GLOCAL_V1.0). A green marked cell means that the content before is already available

3.2 Data models

The SHELTER project is going to integrate local knowledge and existing data of different sources and different nature for the implementation of a Data Driven Platform. These data are provided at a different scale and are created using different tools, formats and approaches by different stakeholders for different purposes and uses. The huge amount of existing data and its heterogeneity make a proper information management strategy for an informed decision-making crucial for the resilience (and vulnerability) assessment on historic areas and cultural heritage. This strategy should link the information at a different scale (building, district, city, region and cross-region), with different hazards (earthquakes, storms, floods, heat waves, wildfires and subsidence), from different domains (cultural heritage, climatological, social, emergency management, etc). Therefore, it will be necessary to design a **Multiscale Multisource Data Model** to structure the relevant information gathered within a complete and unique model, which will support the decision making and tools that will be developed.

The design of this data model is encompassed in Task 1.4 of SHELTER project, whose main objective is to design a data model in order to store the geospatial and semantic data provided by the external sources and the Open Labs and facilitate the management and understanding of information handled by the different actors involved in cultural heritage safeguarding on this project.

SHELTER's multiscale data model will represent the relevant information to characterize the Resilience ID. It is therefore a geospatial data model with information at different scales based on international standards such as CityGML and INSPIRE. The data model will be deployed in a geospatial database and the information will be made accessible through services provided by a GIS server. The multiscale model will be fed and connected with geospatial information from external sources and each of the Open Labs, as well as reference to data included in the Data Lake. The information of the multiscale model will be used from the SHELTER platform tools, mainly Decision Support System (DSS) and Dashboard. The following figure shows the Multiscale Multisource Data Model concept:

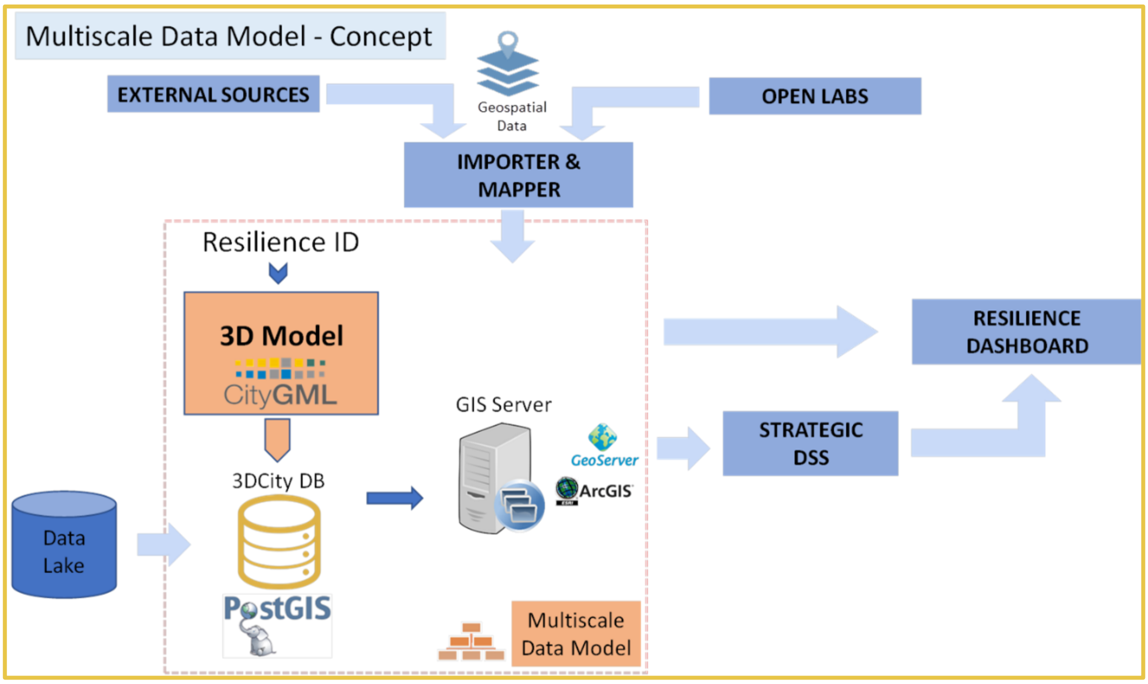


Figure 8 - SHELTER's Multiscale Multisource Data Model concept

Within this task the following work has already been carried out:

- Identification of the general requirements for the Multiscale Multisource Data Model in SHELTER context.
- Definition of the flow of information and the connection of this model within SHELTER overall strategy.
- Analysis of the available data sources.

At the same time, progress is being made on the following objectives:

- Identification of elements, relationships and metadata of data model.
- Obtaining a design of SHELTER data model compatible with the existing data, considering the particularities of the heterogenous information provided by multiple sources of data and the different scales that SHELTER is going to work on.
- Definition of data model considering the main European Standards and programs facilitating the interoperability with models and tools in common use.
- Validate the selection of the standard City GML as a base of the SHELTER data model:
    - customize the data model to the needs of SHELTER project through the inclusion of three application domain extensions for CityGML: Cultural Heritage Domain Extension, Indicators Extension and Dynamic Data Extension.
- Adaptation of the model to be compatible with the concepts within the specific ontologies identified as well as with the best practices.
- Design of the model in a clear way in order to be implemented later in WP5.

The Multiscale Multisource Data Model will take into account the already existing systems, as for example the Sava GIS already developed before the launch of SHELTER (see section 4.6), integrating them in compliance with the principle of interoperability. In this sense the mock-up for CH attributes template developed in the framework of Task 1.2 (see section 0) fulfils both the requirements of local context and the criteria of interoperability, which allow the link with the described data model.

Already existing systems most of the time cannot be restructured and it is the task of data models to foresee the right strategy for their consistent integration. With this regard, the data collection through the DMF and the data example approach (see section 3.1) has been extremely important for the support given to those SHELTER Tasks in charge of developing new technologies, in general, and to Task 1.4, in particular.

4 SHELTER data sources for CH climate and disaster resilience

The Data represents the cornerstone for any innovation in the digital era. SHELTER aims at implementing a seamless digital area with the scale that will enable data exploitation functions like data processing, data analysis and data visualization giving more value to the data generated by the public sector that represents a great European asset.

In the resilience of HA against CC and Natural Hazard events, the data dealing with the cause of the problem are mainly related to meteorological, climate and natural events which can be monitored and measured at global and local scale.

To operationalise the existing heterogeneous data and knowledge, SHELTER aims at making use of the most advanced Information and Communication Technologies (ICT), by means of Earth Observation (EO) data, climate models and local measurements. Such data are in fact fundamental to implement what is called the GLOCAL paradigm (see D6.1_GLOCAL_V1.0), as local data can explain the effect of global climate phenomena. This local data, together with socio-economic information available in the case studies, are worthwhile to give local answers to global problems.

The data and knowledge operationalisation is achieved through diverse methods as:

- a Data Lake for heterogeneous data (satellite imagery, sensor data, geo-environmental and social big data, existing building and disaster databases and crowdsourcing)
- a Multiscale Multisource Data Model, to structure all information from case studies
- a 'Best/next Practices Observatory', that links the portfolio of sustainable and cost-effective solutions for adaptation and reconstruction, governance schemes, co-creation processes blueprints and resilience financing and business models.

The next sections describe the different kinds of data that have been considered useful for SHELTER scope, with particular reference to their importance in monitoring and detecting hazards, mitigating their effects as well as developing disaster resilience.

4.1 Indicators and user requirements

Both the user requirements investigated in the framework of Task 6.1 and the indicators defined in Task 2.2 are of fundamental importance for identifying the set of data that could be exploited for SHELTER objectives.

Indicators

A robust resilience capacity measurement framework requires a set of indicators contributing to quantify the performance of the system as a whole regarding its preparedness and the ability to absorb disturbances, to efficiently respond, and adapt to new conditions. SHELTER methodology will be based on a set of indicators that will quantify the links between multiple dimensions of HA resilience, the connections between different spatial scales and the changes across temporal scales. The system will be developed and implemented in collaboration with stakeholders to lead to development of action plans for enhancing resilience and will have as result a HA Resilience Index and KPIs for resilience monitoring, co-monitoring of the project results (in Open Labs) and benchmarking tool.

As already mentioned in the previous sections, the list of indicators defined in the framework of Task 2.2 has been used in support of the selection process for the data described through the DMF (see section 3.1.1.2). Being the finalization of such a list still in progress at the moment of writing the present document, a further review of the selected indicators will be necessary as foreseen in the roadmap described in section 7.2.2. Meanwhile a short list of indicators, based on the RACER evaluation method, has been produced by Task 2.2 partners and shared within the consortium, in order to start working on the task actions which require such input.

The RACER methodology evaluates indicators according to five criteria: Relevance, Acceptance, Credibility, Easiness and Robustness. Relevance is given if the indicator is closely linked to the objectives to be reached; acceptability is given if the indicator is perceived and used by policy makers and civil society; credibility is measuring the methodological transparency; easiness to compile indicates the possibility to produce readily available data; and robustness indicates high data quality.

Thus, applying the RACER framework allows assessing the general value of scientific tools for their use in policy making and providing an indication on the general properties and quality standards of indicators.

Table 5 and Table 6 show a further selection of indicators extracted from T2.2 short list by taking into account only the indicators to which the higher RACER evaluation rate has been assigned.

Category	Subcategory	Indicator
Hazard characterisation	Rainfall characterisation	Daily maximum precipitation corresponding to the return period T
	Flood hazard	Flood area corresponding to the return period T
		Flood frequency: linked with the return period
	Heat waves characterisation	Daily mean temperature
		Thermal shock [T_{max}-T_{min}]
		Daily sun hours
		Mean relative humidity
		ONRN indicators for heatwave
		Daily humidity cycle shocks [RH(n)-RH(n+1)>25%]
		Relative humidity concentration [nRH>75%]

		Temperature above 35ºC for more than 3 consecutive days
		The daily maximum temperature of more than 5 consecutive days exceeds the average maximum temperature by 5°C, the normal period being 1961–1990
	Storm characterisation	Wind speed
		Lifted index
Receiver characterisation	Pollution	Air quality
	Evolution of the ecological niche	Annual Mean Temperature
		Annual Precipitation

Table 5 – Environmental (meteo-climate, ecological…) indicators extracted from T2.2 short list of indicators (and to which the higher RACER evaluation has been assigned)

Category	Subcategory	Indicator
Receiver characterisation	Demographic vulnerability	% Population below 65 years of age
		% persons 17 years of age or younger
		% Population without sensory, physical, or mental disability
		Percent of female
		Percent of one person household
		Net international migration
	Demographic	Population density (people/5 Km2)
	Communication	Percent population with a telephone
		% Population with access to broadband internet service
Receiver characterisation	Urban characterisation	Average percent perviousness
		Dam capacity
		Average slope
		Average elevation
		Share of the protected lands
		Share of ecological corridors
		Urban growth, avg. annual rate (%)
		Land take
		Land take in hazard area
		Land cover
		Critical facilities in hazard area
		Height above sea level
		Subsidence rate
	Economic	Unemployment rate
	Social capital	Red cross volunteers per 10,000 persons
		# of registered volunteers
		High multiple hazard area
	Exposure	Buildings in hazard area
	Hazard characterisation	Frequency of disaster event
	Building characterisation	# of one floor houses
		Thermal diffusivity
		Solar reflectance

	Soil	Soil aggregate stability
		Soil water capacity
Exposure	Demographic exposure	Population in hazard area
Preparedness	Infrastructure	Psychosocial support facilities per 10,000 persons
		Hospital beds per 10,000 persons
	Building characterisation	% of buildings complying with hazard-resistant building codes and/or standards
Recovery	Damage characterisation	High multiple hazard area

Table 6 – Statistical (demographic, urban, economic…) indicators extracted from T2.2 short list of indicators (and to which the higher RACER evaluation has been assigned)

Both the environmental and the statistical categories of indicators have been analysed in terms of data availability and used as content-related criteria of selection, as explained in section 3.1.1.2.

Another relevant source to be considered as reference for exploring, downloading and analysing indices of observed and modelled climate extremes is the **Climdex** project which offers a list of 27 climate extremes indices. These indices are annual or monthly statistics of modelled or observed climate data. Here one can find descriptions and formulae for each of the indices.

User requirements

In the context of Task 6.1 (particularly Subtask 6.1.2 and 6.1.3) both a bottom-up and a top-down requirement analyses have been performed, in order to establish a user's oriented framework that could guide all SHELTER developments with main focus on WP1, WP3 and WP5 activities (see D6.1_GLOCAL_V1.0).

For what concerns the bottom-up approach, the critical user requirements of end-users have been identified by utilizing '*Use Case Scenarios*' (UCS) as a method to elicit and prioritize the end-users needs within the scope of SHELTER as regards Climate Change Adaptation (CCA), Disaster Risk Management (DRM), and Cultural Heritage Management (CHM). Overall, through the combination of different approaches, a total of 116 user requirements were identified.

The identified User Requirements are structured as follows:

• **General** – includes all UR which are basic for the SHELTER system in relation to the technical functionality as well as in relation to the identified stakeholder roles.

• **Data** – includes databases, important information identified and specific plans like evacuation plan or measurement plan.

• **Analysis** – includes, amongst others, UR dealing with monitoring and early warning systems.

• **Visualization** – summarizes the identified UR to visualize relevant content on digital maps including 3D as well as over time.

• **Crowd** – summarize the UR which are identified to communicate with people in both directions (*e.g.*, warning messenger; reporting system, …).

• **Models** – includes for example foresight and state of the art models (*e.g.* wildfire model, meteorological model, …).

As the five OL's are unique, facing distinctly different hazards and types of CH, the structure of the identified user requirements (general, data, models, etc.) was crucial in establishing some key '*common ground*' between the different OL partners. The outcome derived from the further selection of the identified user requirements (based on the use of prioritization terminology within the questionnaire submitted to the stakeholders - '*must-have*,' '*should have*,' '*nice to have*,' '*not necessary*') is vital for the development of WP1 as well as for technical partners in WP3 and WP5. In particular, the activities performed within WP1 and WP6 are characterized by two-way exchanges, with the former establishing what types of data are relevant and needed for the OL's, as a rough overview, and the latter developing a digital survey as an information-gathering instrument for experts in DRM and CHM, including also relevant questions for WP1 (the complete concept for the digital survey is part of Annex I in D6.1, where the specific elements for WP1 are marked with a cloud in the mind map).

4.2 The Copernicus programme and services

The Copernicus programme is one of the European flagship programmes, providing free and open data and information relying on satellite-based imagery, models and in-situ data. It relies on state-of-the-art models to be used for societal and environmental purposes. The Copernicus programme is a public service designed to respond to policy and public administrations, and foster economic growth in Europe by:

• Supporting public users at local, national, and European level

• Helping Europe to maintain a prominent role in the international context

• Strengthening intermediate users, downstream companies, and value-added service providers.

Initially developed to focus on environment and security – the former name of the Copernicus programme was *Global Monitoring for Environment and Security (GMES)* – the Copernicus programme has developed **six core services providing free data and**

information, enabling applications in a vast variety of fields (*i.e.* agriculture, biodiversity protection, air quality, search and rescue, etc.).

The European Commission (EC) is managing the Copernicus programme and its 3 main components: Space, Services and In-situ components. The high-level structure of the Copernicus programme is presented in the figure below:

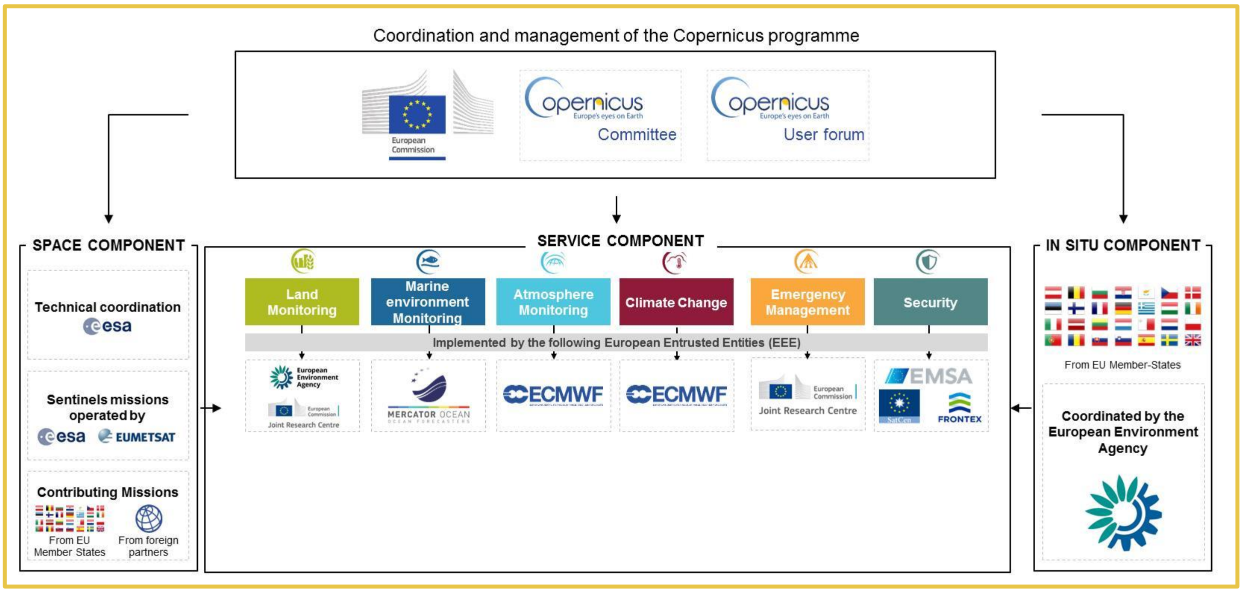


Figure 9 – High level structure of the Copernicus programme

The Copernicus Space Component deals with the procurement, launch, operation and the distribution of Sentinels data and of contributing missions' data. The technical coordination and procurement for the Sentinels fleet are led by the European Space Agency (ESA) and operated by collaboration between ESA and EUMETSAT. This element also includes the procurement of the overall space infrastructure, including satellite design, satellite manufacturing (procurement to the industry), satellite launches and ground infrastructure manufacturing (procurement to the industry). Finally, ESA is also in charge of acquisition, storage and distribution of the Sentinels data via the ESA Scientific Hub platform. As a transnational space agency collaborating with all the European national space agencies, ESA has access to several national EO programmes' data, including the archives of such programmes. This additional data source is called "contributing missions" and provides, for registered users, access to a wide range of commercial (*i.e.* Worldview, SPOT, TerraSAR, Radarsat 2, etc.) and civilian (*i.e.* Landsat, COSMO-SkyMed, RISAT, etc.) EO data sources. These data sources offer in some cases higher spatial resolution than the Sentinels spacecraft, to support the development of specific information products provided by Copernicus core services. However, the access to contributing missions is based on restrictions and so not fully open to everyone. For obvious reasons, high and very high-resolution imagery is only open to a restricted list of authorized users in the field of security and emergency.

In the past decade, increased availability of high-resolution satellite sensors has contributed to great improvements in hazard detection and mapping. Satellite and drone images can facilitate quick, large-scale assessment of the impact of a disaster by comparing pre- and post- event images of damaged buildings and infrastructure for example.

Previous approaches to remote sensing, which used coarse data and relied on physical models built and interpreted by experts, rarely allowed global coverage, frequent updates, or the fusion of data from multiple platforms. Improvements in the velocity, volume, and variety of satellite imagery data, along with automated methods for processing and aggregating data, have been a boon for exposure mapping.

Assessing exposure and vulnerability to hazards is now possible through new satellite imagery products available at relatively low or no cost, thus allowing the development of new algorithms for mapping purposes.

As explained above a strong limitation to the use of satellite imagery for increasing CH climate and disaster resilience lies in the relatively low spatial resolution when compared to the small scale characterizing certain case studies (*e.g.* Ravenna).

In order to overcome this issue SHELTER combines satellite imagery with other types of data sources, such as IoT, crowdsourcing, predictive and statistical models, etc. Through this combination of heterogeneous data, satellite imagery can heavily support resilience analytics, particularly for:

- Descriptive analytics: satellite imagery can be combined with social media data for assessing damages from a disaster (*e.g.* early detection of a flood)

- Predictive analytics: satellite data are central to enabling granular, early and accurate weather forecasts and can increasingly predict both sudden and slow-onset disasters.

Among the numerous EO data providers, the Copernicus program makes available an amount of data never seen before including data gathered by Sentinel sensors family and SHELTER aims at integrating, connecting and exploiting all the Copernicus data and products related to tangible, intangible and natural cultural heritage preservation and management.

The '*Copernicus services in support to Cultural Heritage report*' has been used as reference to identify the main EO-based Copernicus products useful for SHELTER

The Copernicus In-situ component offers access to observation from the ground, sea and airborne sensors but also licensed reference and ancillary data licensed; in-situ data are not freely available for Copernicus users. The in-situ component supports the space component in offering access to sustainable and reliable data to produce, validate and calibrate Copernicus products for the services component. The In-situ component is implemented in two tiers:

• At the level of the service: each core service is in charge of daily operation and ingestion of specific in-situ data of interest per thematic (marine service, land monitoring, etc.) to offer valuable products for their end-users. This means that specific sources of in-situ data are tailored for each core service.

• At the programme level: the European Environment Agency (EEA) manages the cross-cutting service offering general in-situ data accessible through specific agreements with data providers/networks at programme level.

The Copernicus Services component aims to deliver data and products freely available for a wide variety of users. These services integrate data from the Space and In-situ components, together with state-of-the-art models, in order to offer Copernicus products tailored to the needs of specific end-users. To better reach end-users, six different core services are developed in different areas:

• Copernicus Land Monitoring Service (CLMS):

CLMS is split into three components: the Global component (managed by the DG Joint Research Centre – JRC), the Pan- European component, and the Local component (managed by the European Environment Agency). It has the aim to provide geographical information on land cover, land use, land cover-use changes over the years, vegetation state and the water cycle. It is mostly used for forest management, water management, agriculture or food security.

• Copernicus Marine Environment Monitoring Service (CMEMS):

CMEMS is managed by Mercator Océan International, with the aim to provide regular and systematic information about the physical state and dynamics of the ocean and marine ecosystems. Its products cover the global oceans and the European regional seas, through the provision of observations and forecasts. It is mostly used for ship routing services, offshore operations or aquaculture

• Copernicus Atmosphere Monitoring Service (CAMS):

CAMS is managed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and has the aim to continuously monitor the composition of the Earth's atmosphere at global and regional scales through the provision of near-real time data and forecasts products. It is mostly used for health, renewable energy, or climatology issues.

• Copernicus Climate Change Services (C3S):

C3S is managed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and has the aim to respond to changes in the environment and society associated with climate change, through the provision of information for monitoring and predicting climate change and help to support adaptation and mitigation strategies. It is mostly used for climate, weather and renewable energy monitoring.

• Copernicus Emergency Management Service (EMS):

CEMS encompasses two components: the early warning component is managed by the DG Joint Research Centre (JRC) and the mapping component is managed by DG for European Civil Protection and Humanitarian Action (DG ECHO). The aim of the former is to deliver warnings and risk assessments of floods and forest fires, while the mapping service provides map and geo-information products for all types of natural and man-made disasters. CEMS intervenes both at European and global levels.

• Copernicus Security Service (CSS):

The Copernicus Security service aims at improving crisis prevention, preparedness and response in three domains: border surveillance (managed by FRONTEX), maritime surveillance (managed by EMSA) and support to European Union (EU) External Action (managed by EU SatCen). It is mostly used to support related European Union policies by providing information in response to the security challenges Europe is facing.

The Copernicus services were designed to respond to very specific needs of the European society, targeting specifically public authorities but also research and scientific communities. Nevertheless, the quantity and quality of the data and products offered by services also respond to commercial end-user needs. In this context, most of the products provided for free and openly accessible for everyone were designed with an objective of ensuring the European downstream industry.

In the following sections some examples of datasets selected for SHELTER purposes from Copernicus service and products are presented and described.

4.2.1 Copernicus Space Component in SHELTER

Within SHELTER the main satellite data from Copernicus that will be considered are the Sentinel 1 (S1) Synthetic Aperture Radar (SAR) grd and the Sentinel 2 (S2).

The **Sentinel 1** is a mission operated by ESA composed of two flying satellites. SAR is the type of sensor on board of the satellites and it stands for Synthetic Aperture Radar. The frequency of data acquisitions also known as revisit time is 6 days and the ground resolution of the images spans from 6 to 40 meters. The data are available since 2014.

The main feature of this sensor is that as a radar it operates in the electromagnetic spectrum range of the microwaves (Band C) and it is able to detect surface characteristics even in presence of clouds.

The scope of this mission is the monitoring of Arctic sea-ice extent, routine sea-ice mapping, surveillance of the marine environment, including oil-spill monitoring and ship detection for maritime security, monitoring land-surface for motion risks, mapping for forest, water and soil management and mapping to support humanitarian aid and crisis situations.

In SHELTER the S1 data are used for flood mapping service and in wildfire danger assessments:

- Sentinel 1 surface soil moisture (DMF ID: *32007*): Surface Soil Moisture (SSM) is the relative water content of the top few centimeters soil, describing how wet or dry the soil is in its topmost layer, expressed in percent saturation. It is measured by satellite radar sensors and allows insights in local precipitation impacts and soil conditions. Knowledge on the dynamics of soil moisture is important in the understanding of processes in many environmental and socio-economic fields, *e.g.*, its impact on vegetation vitality, crop yield, droughts or exposure to flood threats. Soil Moisture is recognized as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS). Within SHELTER this kind of data is particularly useful for Baixa-Limia Serra OL, being this area prone to wildfire hazard.

- Sentinel 1 SAR grd images: this satellite data has been selected as highly recommended for detecting flood extent information after a flood event. This information is in fact very important for damage assessment and risk management, and benefits to rescuers during flooding. As a matter of fact, one of the most important problems associated with flood monitoring is the difficulty to determine the extent of the flood area as even a dense network of observations cannot provide such information. SAR measurements from space are independent of daytime and weather conditions and can provide valuable information in this sense. It considers the increasing amount of backscattering in case of flooding on different land use classes: bare soil, vegetated areas and urban areas. Therefore, different SAR-based mapping algorithms are used according to the identified proper land cover classes; the different classes are used to detect floods and images are split in different areas. One of the methodologies that can be applied to implement a flood mask custom procedure comes from the United Nations, with the UN-SPIDER Recommended Practice on radar-based flood mapping. The practice shows the use of ESA SNAP software for pre-processing and processing of SAR imagery using a threshold method for deriving the flood extent. Google Earth is used to visualize the results of the image processing. This practice can be applied globally and has been used successfully for floods in Australia, Africa and Asia (see the '*UN recommended practice for flood mapping'*).

  At the time of writing this deliverable, new methodologies customized for SHELTER are under development (Task 3.2 **– Rapid damage assessment technologies**). Sentinel 1 SAR grd images will be exploited to obtain water masks of flooded areas so that they can be used as input (calibration maps) of the flood hazard models implemented in the project, in particular for its quantitative information on the extension of the flooded area and on the potential damages.

The **Sentinel 2** mission comprises a constellation of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other. It aims at monitoring variability in land surface conditions, and its wide swath width (290 km) and high revisit time (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes) will support monitoring of Earth's surface changes.

The scope of this mission is vegetation monitoring, urban monitoring and agriculture applications.

In SHELTER, the S2 data are used for vegetation monitoring, burned areas and drought.

The following products have been considered useful for the project scope:

- Sentinel2 true colour (DMF ID: *32002*): the true color composite is a widely used Earth observation product for displaying satellite imagery, as it can be used in comparison with other many satellite products to detect changes in the observed territory, due to the impact of different hazards (flood, wildfire, earthquake…).

- Sentinel2 Normalised Difference Vegetation Index – NDVI (DMF ID: *32008*), which is an indicator of the greenness of the biomes and is widely used by the bio-geophysical community to monitor the vegetation state and disturbances to address a large range of applications, including forestry, agriculture, food security, water management. In the context of SHELTER this product is particularly relevant for the Baixa-Limia Serra OL, which is exposed to wildfire hazard: by means of the analysis of its temporal evolution in fact, the NDVI can be used as an indicator of water stress, thus helping in estimating forest fire danger (prevention DRM phase). Moreover, this index can be very useful to measure post-disaster vegetation response, through the comparison between pre- and post-fire NDVI values (response DRM phase).

4.2.2 Copernicus Service Component in SHELTER

Within SHELTER the main data services from Copernicus that will be considered are the CLMS, the C3S, the CAMS and the CEMS.

CLMS

With regard to this service, the Corine Land Cover (CLC) is a European programme, coordinated by the EEA, providing consistent information on land cover and land cover changes across Europe. CLC products are based on the photo interpretation of satellite images by the national teams of the participating countries – the EEA member or cooperating countries. The resulting national land cover inventories are further integrated into a seamless land cover map of Europe. The CLC is provided for 1990, 2000, 2006, 2012, and 2018. This vector-based dataset includes 44 land cover and land use classes. The time-series also includes a land-change layer, highlighting changes in land cover and land-use.

The following two products have been selected as relevant for SHELTER, particularly for the Baixa-Limia Serra OL, as highlighted in the related user requirement '*vegetation data*' (see D6.1_GLOCAL_V1.0, Figure 16):

- the CLC 2018, which refers to land cover / land use status of year 2018 (DMF ID: *73003/75053*)
- the Corine Land Cover Change (CHA) 2012 – 2018 (DMF ID: *73004/75054*)

Land cover distribution is in fact one of the factors that influence fire behavior and its consequences in the landscape.

Moreover, the following Digital Elevation Model (DEM) product has been selected for its high spatial resolution (25 meters) and improved vertical accuracy with respect to the previous version:

- EU-DEM v1.1 (DMF ID: *72048/73006/74012/75056/76032*): this is a digital surface model (DSM) of EEA39 countries representing the first surface as illuminated by the sensors. It is available in Geotiff format, thus allowing its use through Geographic Information System (GIS) tools and it is represented by a contiguous dataset divided into 1000 x 1000 km tiles, with the tiles that are grouped in big regions. Common uses of DEMs include: extracting terrain parameters for geomorphology, modeling water flow for hydrology or mass movement, modeling soil wetness. Among the user requirements identified in SHELTER, this dataset is well suited with '*site characteristics (morphology, geology, urban functions, etc.)*' (see D6.1_GLOCAL_V1.0, Figure 16).

C3S

Within this service, the ERA5 products represent the fifth generation ECMWF atmospheric reanalysis of the global climate. Reanalysis datasets are important because they combine model data with observations from across the world to create a globally complete and consistent dataset giving a better overview of weather and climate conditions. This is particularly important for areas where observational data is either sparse or non-existent. ERA5 is particularly powerful because of its improved model accuracy and higher geographical and temporal resolution than its predecessors. Starting from this dataset climatologists and meteorologists can perform better forecast analysis and downscaling modelling.

The following ERA5 products from the C3S have been selected as relevant for SHELTER scope:

- ERA5 daily average temperature of air at 2m above surface (DMF ID: *32005*)

- ERA5 daily maximum temperature of air at 2m above the surface (DMF ID: *32006*), which, together with the previous one, can represent the basis for calculating some of the indicators identified in the framework of Task 2.2 (see section 3.1.1.2 and Table 5), such as the *Temperature above 35ºC for more than 3 consecutive days* and thermal anomalies in general.

- UERRA regional reanalysis for Europe on single levels (DMF ID: *73005*), which provides, among other variables, an estimation of the relative humidity at 2 meters above the model topography. This parameter is particularly relevant for some of the identified indicators (see section 3.1.1.2 and Table 5) and meets Seferihisar OL requirements.

- ERA5-Land (DMF ID: *73007*): this is a reanalysis dataset providing a consistent view of the evolution of land variables over several decades at an enhanced resolution compared to ERA5. It has been produced by replaying the land component of the ECMWF ERA5 climate reanalysis and it shows both a spatial and temporal resolution higher than the previous ERA5 products and a quite large temporal coverage, making it more useful for applications at local scale as those related to SHELTER OLs. As the previous ones, it is particularly useful for the Seferihisar OL scope. Among the numerous variables of which ERA5-Land is composed, those grouped under 'Temperature', 'Soil water', 'Radiation and heat', 'Evaporation and run-off' and 'Vegetation' categories are of major interest as for their relation with heat waves hazard.

CAMS

Among the products available from this Copernicus service the following one has been selected as considered useful particularly for Ravenna OL:

- Daily air quality analyses and forecasts for Europe (DMF ID: *72047*)

  This dataset provides daily air quality analyses and forecasts for the European domain. The production is based on an ensemble of nine air quality forecasting systems across Europe. The analysis combines model data with observations provided by the EEA. In parallel, air quality forecasts are produced once a day for the next four days.

Air quality is one of the indicators that has been marked in Task 2.2 with the highest rating (see section 4.1). It is a well-known fact indeed that CH sites can deteriorate, since anthropogenic activities generate air pollution that can be dangerous to buildings and the environment (Sablier, M., Garrigues, P. (2014) [2]). In this sense such

Copernicus product can support those SHELTER OLs that are included in the Tangible Heritage category (*i.e.* Ravenna, Seferihisar and Dordrecht)

CEMS

Among the different products that this Copernicus service provides, both the Risk and Recovery Mapping (RRM) and the Rapid Mapping (RM) come with a list of EMS activations that can be triggered through a Service Request Form (available on the Copernicus EMS website). Such emergency activations can be used as reference for selecting a list of events described in terms of time range and area coverage, thus allowing to retrieve the proper subset of the dataset in use (*e.g.* Sentinel 1 SAR grd images in case of flood extent detection). With the same purpose this service should be taken into account for areas subject to wildfire hazard.

4.3 IoT data

Many definitions of the "Internet of Things" exist. One of these is given below.

The IoT (Internet of Things) is a network that intelligently identifies, locates, tracks, monitors, and manages objects through radio frequency identification (RFID) or other Unique Identifier systems like infrared sensors, global positioning systems, laser scanners, and other information-sensing devices according to defined agreements or communication protocols.

This integrated network includes a powerful central computer cluster that can integrate people, machines, equipment and infrastructure in the network and manage and control them in real-time.

In the case of SHELTER, we understand that the IoT addresses locally installed sensors which are in turn connected to databases in servers. These servers can be installed locally, in the region, the nation or even internationally in the cloud.

IoT plays a very important role in complementing and/or enhancing the accuracy and efficiency of monitoring and early-warning systems for hazards prevention. It is in fact very useful when it comes to filling the spatial resolution gap which can often represent an issue when exclusively satellite data is utilized (see section 4.2).

Moreover, data derived from IoT can compensate for the lack of information frequently related to other data sources.

Finally the IoT provides a good terminal extension technology and equipment for early-warning information release. Once this information is processed, SHELTER or other Risk management Systems can transmit hazards early-warning information through mobile phones, short messages, wireless early-warning broadcasting networks, satellite regional communication broadcasting equipment and so on, thus shortening the release time of information and reducing the risk of hazards.

On the other hand, IoT data sources are not exempt from weaknesses. Various factors may cause the IoT monitoring system unreliable: first, security and unified standards are not always guaranteed; second, as a consequence of the local scale characterizing IoT data, the languages used for the related metadata, instruments instructions and technical definitions are not homogeneous; finally the quality criteria are often unharmonised. By combining many different types of data sources (*e.g.* the ERA5 products described in the previous section), SHELTER aims at compensating the weak points of each source while taking advantage of their strengths.

In order to collect the relevant information on IoT sensors and the related data already existing and available from SHELTER case studies, a survey has been performed with the support of SHELTER technical partners showing a high expertise in the IoT field (RED and POLITO – see Annex IV). Such survey has been submitted to the local stakeholders through a Level 3 interaction (*i.e.* requests raised by technical partners are forwarded to the Open Lab Task Working Group, so that they can interact on this with the targeted stakeholders). Most of the Open Labs have responded to this survey by completing the Data Mapping Form with information regarding the locally installed sensors and the portals where the sensor data are stored.

Most of the local sensors are installed in weather stations or stations monitoring water flow, quality and levels. Through gateways, these sensors in turn transmit their data to servers using one or other data transmission protocol. The data are stored in these servers in databases (SQL or other) from where these data can be accessed and extracted, in most cases via web portals. For automatic extraction of these data via API's, it will be necessary to request authorization to the owners of these data. The use of these data will most likely be governed by additional agreements with the owners. In addition, API's may in certain cases not yet exist and have then to be developed.

This information from the Open Labs has been subsequently analyzed by the technical partners breaking down the information from the weather and other monitoring stations into individual sensor and sensor data. The Data Mapping Form has been completed with the results of this analysis (see sheet *P_16_RED* of the DMF).

The type of sensors installed are generally monitoring temperature, relative humidity, rainfall, wind (speed and direction), radiation (total and visible range), barometric pressure, water level, water flow, water turbidity sensors. Through data storage and processing in the servers, instantaneous, average, minimum and maximum values over days, weeks, months, years can be extracted. This varies obviously from site to site. Figure 10 shows an example of the information collected for the Baixa-Limia Serra Open Lab (see **Annex V** for a detailed overview on the information received):

Estación Entrimo. Entrimo (OU)

Pluviómetro de Cazoletas

Fabricante: R. M. Young Modelo: 52202/52203 Características: • Precipitación + info	

Sonda de Temperatura y Humedad

Fabricante: Campbell Modelo: HMP45AC Características: • Temperatura del aire , incertidumbre: ± 0.2 °C • Humedad Relativa , incertidumbre: ± 3 % + info	

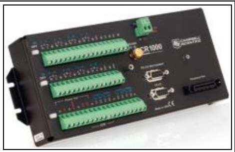
Piranómetro

Fabricante: Kipp&Zonen Modelo: CMP-3 Características: • Radiación solar global + info	

Sonda de Temperatura de Superficie/Suelo

Fabricante: Campbell Modelo: T-107 Características: • Temperatura del superficie + info	

Datalogger

Fabricante: Campbell Modelo: CR1000 + info	

Pluviómetro de Cazoletas

Fabricante: Lambrecht Modelo: 00.15189.002000 + info	

Figure 10 – Information on IoT sensors already installed in Baixa-Limia Serra Open Lab – Xunta de Galicia

With respect to the new (not existing) sensors that are going to be developed and installed within SHELTER, workshops need to be organized with the Open Labs and the different experts in risk management and reconstruction to define the type of sensors to be added to the existing network. A preliminary proposal for the different Open Labs was included in the Description of Action. Local sensors for structural monitoring (such as extensometers, accelerometers), for humidity monitoring in soil and walls, for pluviometry and piezometers are planned to be installed.

At the time of writing the present document, there are no new foreseen sensors for the following OLs: Sava River Basin, Baixa-Limia Serra and Dordrecht. For these OLs, data will be gathered through the sensors already connected to the existing networks.

In the case of Seferihisar OL, the discussion regarding if new sensors are necessary is on-going. The possibility to make the assessment through modelling is under evaluation, as it would give a more immediate result. The decision will be made soon.

For Ravenna OL, a selection of new sensors to be installed is ongoing.

If at all feasible, these sensors should be integrated in the existing networks as much as possible. They could make use of a backend service (gateway) to collect the related data and of the same transmission protocols and servers, or they could be set in a way that data is sent directly to the IoT module developed within WP5 (POLITO). If this is not possible or desirable, then individual monitoring systems need to be set up including gateways and servers. The choice of the sensors will depend not only on the functionality, but also on the availability of local support and the compatibility with the gateways and servers to be used.

Annex IV contains the survey performed for collecting information about existing IoT sensors and the related data, as well as future sensors, with specific questions on communication protocols, monitoring network and parameters to be measured, in order to start defining the minimum requirements and the interoperability/protocol standards through the outlining of the group of product that fulfil the needs of SHELTER and beyond.

4.4 Crowdsourcing and Social Media data

It is known that both crowdsourcing and social media data are contributed by the public. While social media data are unstructured, ad-hoc crowdsourcing approaches designed for a distributed data collection can be more structured to ease the data aggregation and analysis (Stefanidis, A. *et al.*, 2013 [3]).

Platforms based on this kind of data can be developed to improve the disaster response and resource allocation based on real-time reports from disaster victims.

In particular, a number of authors have shown that real time analytics based on social media data provide good opportunities to detect and monitor events automatically (Middleton, S.E. *et al.*, 2014 [4], Nguyen, Duc T. and Jung, Jai E., 2017 [5]). Text messages are the basic source of analysis. Visual analytics through social media data facilitate spatio-temporal analysis and create a spatial decision support environment that assists in evacuation planning and disaster management (Chae, J. *et al.*, 2014 [6]). Given that social media does not rely only on text messages and provides more useful information through images and videos posted by users, visual analytics and image/video-based analysis are becoming more important in extracting the key information from social media posts.

Active social media and crowdsourcing platforms will be developed in the framework of SHELTER Task 3.5 **– Crowdsourcing solutions for citizens engagement in preparedness and response**. Within this task intelligent conversational tools (chatbot) and event detection algorithms for social media will be used both in preparedness and in response phase, thus to improve early warning systems, focusing on vulnerable groups, especially important in historic rural and urban areas such as elderly people and immigrants. Crowdsourcing solutions are in fact able to deliver as well as to retrieve multimedia geolocated contents from people's smartphones, targeting citizens living in the surrounding of HA. At the same time, social media data engines are able to fetch social media data related to HA in real time and automatically classify the content using advanced text analytics, image processing and deep learning algorithms.

4.4.1 Social Media Analysis

Social networks have become one of the most common tools to spread information around the world since, in a matter of seconds, every new post becomes globally available and easily accessible. This provides an invaluable source of real-time knowledge that can be immediately analysed to extract useful information.

While many different platforms offer similar capabilities in terms of research tools, Twitter probably presents the best balance between immediacy and information content, thanks to its concise format. Moreover, exploiting the streaming API provided by the Twitter developer platform, it is possible to filter real-time tweets specifying the language and a set of keywords to be found in the text. This allows for quick and efficient retrieval of only a small subset of posts, minimizing the processing effort and maximizing the amount of relevant data.

A Data Ingestion module will be responsible for tracking a set of different hazards that can occur in Heritage Areas using keywords in different languages and using parallel streams, one for each language-hazard pair. The retried data will be automatically classified by a set of machine learning models. First, tweets not containing useful information will be discarded by a binary classifier. Subsequently, a hazard type and an information type will be assigned to tweets whose information content is considered relevant. To do so, a predefined information taxonomy will be used, which is reported in Table 7:

Hazard type	Information type
▫ Fire ▫ Flood ▫ Storm ▫ Weather anomaly *(e.g. heatwave, frost)* ▫ Earthquake ▫ Landslide ▫ Avalanche ▫ Subsidence ▫ Collapse *(buildings and infrastructures)* ▫ Pandemic *(disease outbreaks)* ▫ Terrorism *(terrorist attacks, shootings)* ▫ Accident (explosions, crashes)	▫ Caution and advice ▫ Donations and volunteering ▫ Infrastructure damage ▫ People affected ▫ People injured or dead ▫ People missing and found ▫ Sympathy ▫ Other useful information

Table 7 - Preliminary definitions of the classification outputs for the text content. Note that the provided definitions are subject to change, as the development of the modules progresses.

Such a taxonomy will be finalized in T3.2. Next, a *Named Entity Recognition* module will try to extract named entities, which can be real-world persons, locations, organisations or objects that can be denoted with a specific name. Together with a recognition stage, the found entities will be possibly linked to existing structured knowledge bases for disambiguation and retrieval of additional content such as geographical coordinates of the locations. Entities will be grouped into four major categories, namely *people, infrastructures* (*e.g.* roads, buildings, cultural heritage sites), *geography* (*e.g.* countries, regions, municipalities) and *time* (dates, intervals).

Additional multimedia content provided by the tweet, such as pictures and videos, will be analysed in parallel by a *Multimedia Content Classification* module that will provide additional labelling using Deep Learning models and following a similar – but simplified taxonomy, which is still in definition.

Finally, the information extracted from text and media will be merged by the Event Detection module, which will provide automated real-time detection of ongoing events.

The output of the classification and event detection modules will be persisted on a relational database with GIS features and periodical backups.

Upon authentication through the *AA Server*, these outputs will be exposed through an *API Gateway*, providing endpoints for labelled data and detected hazards in *JSON* format to be displayed in the *Data Resilience Dashboard.*

For each tweet, the classification endpoint will provide details such as the following points:

- tweet ID: unique ID of the tweet
- timestamp: datetime of the tweet
- text: raw textual content of the tweet
- language: BCP 47 language identifier

- hashtags: list of tags present in the text
- entities: list of extracted and linked named entities
- impact on entities: how they are affected
- inferred location: tweet location, if deductible or provided
- inferred time: timestamp inferred from text
- informative: Boolean
- hazard type: which category
- information type: what kind of information

Through the event endpoint, it will be possible to retrieve the following information:

- A *time interval,* estimating when the event started and finished, defined as a timestamp tuple
- A *location,* estimating where the event took place, indicated as bounding box of the affected area
- A *type* describing which hazard category it belongs to
- A *status*, indicating whether the event has been confirmed by a responder
- Impact data, indicating the estimated impacts in terms of affected people and/or infrastructures

The definition of the aforementioned data is preliminary and the final definition will be performed in T3.5 and it will depend on the results that will be achieved by the machine learning algorithms, which will be trained on past labelled data available from previous studies.

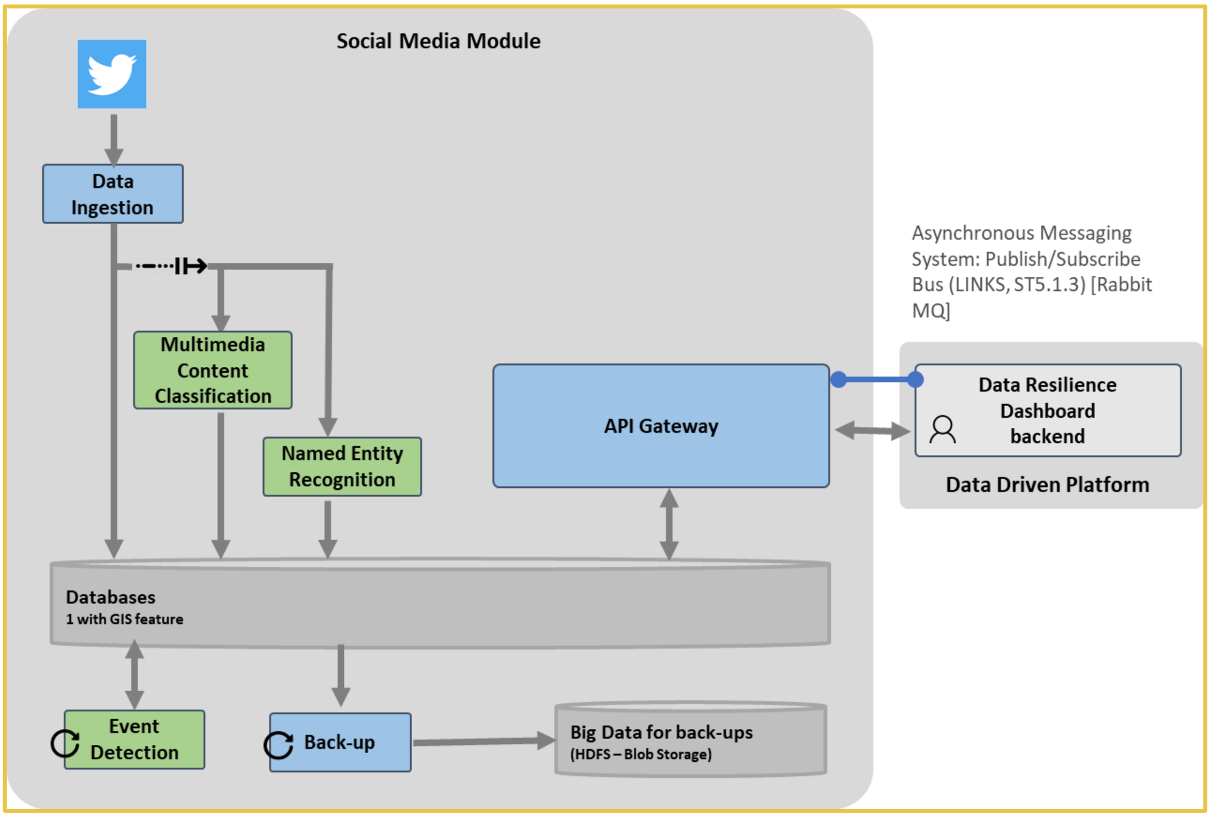


Figure 11 - Social Media Data Engine architecture schema

4.4.2 Crowdsourcing solution

Crowdsourcing is a cheap yet effective way of collecting precious, up-to-date information of several kinds. For some use-cases, it is enough to collect such data using automated social media crawlers, while in other cases, a more structured solution is required to obtain more reliable data, requiring reaching and engaging specific groups of users.

Several channels can be exploited to gather user-generated data, *e.g.* online websites, mobile applications. In SHELTER, the selected approach is the implementation of a Chatbot, an easy-to-use conversational tool which has become more and more popular in the last decade.

The chatbot has been chosen over other kinds of solutions given its flexibility, ease of deployment and maintainability, mobile platform compatibility, high user engagement potential, and given user experience and usability considerations.

Deploying and updating a chatbot on a popular chat application like Telegram is a very straightforward process compared, for instance, to submitting an app to a mobile app store. Furthermore, the chatbot can be deployed on a wide range of compatible clients on different operating systems and devices, thus allowing to reach a large user base and to have a smooth and uniform user experience across all supported clients.

The main purpose of the SHELTER Chatbot is the gathering of user reports, which consists of geo-located contents with multimedia, text and/or more structured data that can be collected answering direct questions or through bot dialogs. Bot dialogs are sequences of chatbot-generated User Interface (UI) controls that allow the user to input relevant information in a guided fashion.

Figure 12 shows an example of reporting using a chatbot:

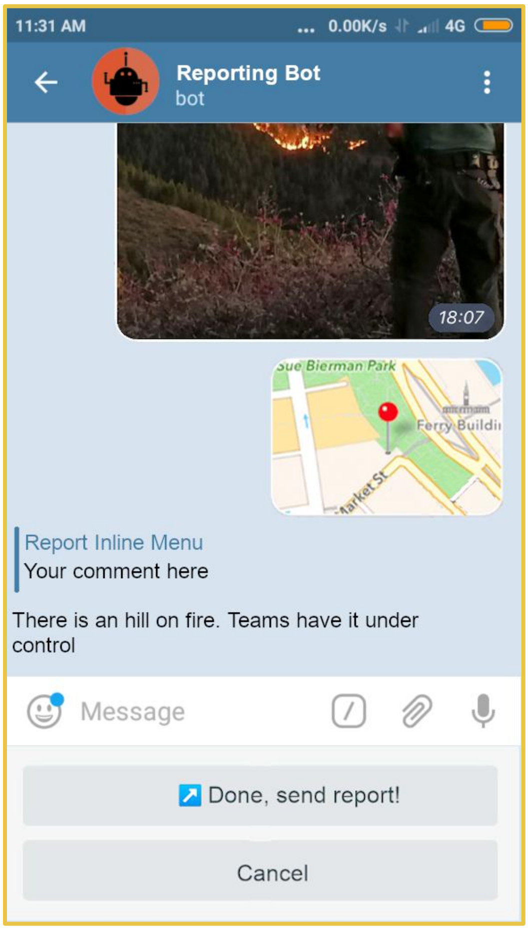


Figure 12 - An example of bot collecting user-generated content on Telegram. It includes a picture, a location and some free text. A custom keyboard and other on-screen buttons allow the user to control the data flow and decide which contents to share.

The informative contents that the SHELTER Chatbot will collect are:

- Free text
- Multimedia (*e.g.* pictures, short videos or audio files)
- User location and Report location (indicated by the user using a map)
- Structured data

A report will be considered valid if at least one data between multimedia and free text is available, together with the User location. The Report location is another optional location that the chatbot will collect to allow the reporter to safely provide data without approaching unsafe locations. To handle these situations, which are common during emergencies, the bot will allow submitting an additional location selected from the map.

Structured data will be domain-specific content related to HA that can include quantitative or qualitative (selected from a closed set) information. The structure of this kind of data will be defined and agreed with the Open Labs, or otherwise derived from a past project (*i.e.* I-REACT). The information provided by structured data is more concise and direct, allowing a quicker classification of the report, while the multimedia and free text content will require to be post-processed and analysed with the aid of Machine Learning (ML) services.

The figure below shows the data flow for the SHELTER crowdsourcing Chatbot.

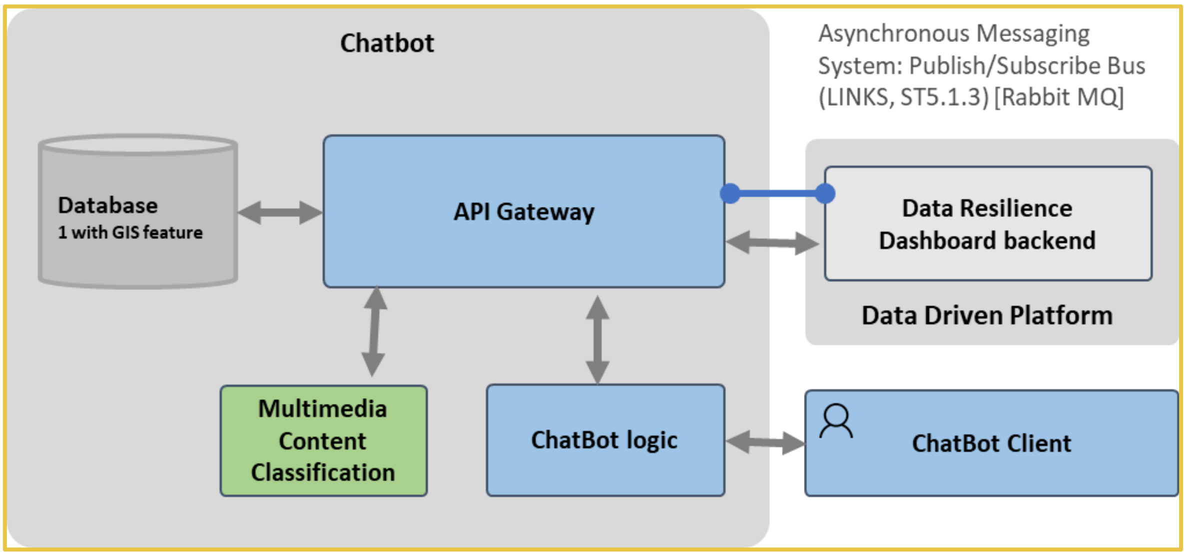


Figure 13 - Data Flow of the chatbot and related components

4.5 Socio-economic data sources

There are constant changes on the earth's surface. Their cause is both natural processes and human activity. We observe around us how people are currently affecting the landscape more powerfully and faster than nature itself. Nevertheless, the extremes in the behaviour of nature have a significant effect on the development of civilization and the current population. They are extremely fast natural processes that have a source in the atmosphere, water, on the earth's surface, in the earth's crust, and even in the earth's mantle. They are independent of the activities of mankind, or, as we argue, man influences them rather indirectly. If one cannot prevent such situations, one must at least limit their effects.

Each of these events has socio-economic impacts on the society / state in which the event occurs and affects the global socio-economic situation to a greater or lesser extent. A reasonably controlled state records these changes in the form of a summary of historical events in writing, photographic record, digital record, enumeration of damages, etc., which subsequently affect economic, social, political life and signed-on citizens' views. Now, there are changes in the economic policy of individual states, but in principle this occurs through the implemented economic policy and legislative framework. At present, we see and perceive how difficult it is for the political representatives of all countries to enforce any reforms, mainly because they have the greatest impact on the socio – economic development of the countries. The rate of inflation, more precisely price stability, and the unemployment rate are among the variables that are in the forefront of interest not only for politicians, but especially for citizens. The main goal of all stakeholders is to find, despite significant losses due to natural disasters, economic stability.

With a detailed look at the objectives and their breakdown, it is known that, due to the conflicting objectives, it is a fundamental problem to ensure a balance between economic policy and the risks arising from real natural disasters. In practice, we see the efforts of national governments to ensure macroeconomic stabilization and to create crisis models in the event of natural disasters, which are based mainly on statistical data and show inconsistencies in fiscal and monetary measures. Thus, these models become theoretical models and their starting points are more likely to detach themselves from the real world.

However, it must be emphasized, in the context of modern and advanced economies, that these crisis models primarily use socio-economic data, statistics to demonstrate the reasons and evidence for deciding on potential national economic policy measures to prevent societal decline in each area. In the framework of SHELTER Task 1.1, the main importance of collecting useful data is their identification, classification, and subsequent evaluation. All data must meet the criteria in terms of quality, readiness, relevance, usability, and availability in the medium / long term, based on historical data, geographical representativeness and thematic significance.

The need of socio-economic data in support of SHELTER scope is clearly demonstrated, in addition to the above mentioned concepts, also through the definition of the list of indicators performed in the framework of Task 2.2 (see section 3.1.1.2 and Table 5) which includes, among the others, a high number of indicators derived from statistical data.

4.5.1 Identification of data sources and knowledge

The data sources considered relevant in terms of socio-economic aspects have been described through the DMF under the sheet dedicated to the technological partner in charge of this task (see *P_7_UMAS* of the DMF). All the identified data sources have been temporarily marked as '*associated with statistical indicators*'. Once the indicators defined in Task 2.2 will be validated, each data source will be more precisely correlated to the respective indicator.

Global, international and local data sources have been identified. The first two are mainly represented by central EU statistical databases and data sources from non-SHELTER countries, while the latter derive from statistical offices of individual EU countries and Open Lab's statistical offices.

The Eurostat database (see ID *11009* of the DMF, *P_7_UMAS* sheet) provides statistical information to the Institutions of the European Union and promotes the harmonization of statistical methods across its member states. The advantage of using this data source within SHELTER lies in the possibility to link the provided statistical data to certain environmental parameters for example. Eurostat, in fact, provides a range of statistics and accounts about the state of the environment and the drivers, pressures and impacts of our societies on the environment (see the '*Eurostat environment overview*'). Thus, information is available, among the others, about: air emissions, biodiversity, environmental protection, water, etc.

Other relevant data sources are available through international providers, such as the Munich RE (Munich Reinsurance Company, see ID *11024* of the DMF, *P_7_UMAS* sheet) a reinsurance company based in Munich, Germany, which provides useful reports and data synthesis.

Also, local statistical sources have been identified for the five SHELTER Open Labs. Following some examples:

1. **Ravenna**:
    - Istat (the Italian National Institute of Statistics, see ID *11005* of the DMF, *P_7_UMAS* sheet) is the main producer of official statistics in Italy. Its activities include the census of population, economic censuses and a number of social, economic and environmental surveys and analysis. Istat is an active member of the European Statistical System, coordinated by Eurostat.
2. **Seferihisar**:
    - TurkStat (Turkish Statistical Institute, see ID *11001* of the DMF, *P_7_UMAS* sheet) is the Turkish government agency commissioned with producing official statistics on Turkey, its population, resources, economy, society, and culture.
    - Ministry of Culture and Tourism (Republic of Turkey, see ID *11012* of the DMF, *P_7_UMAS* sheet) is a government ministry of the Republic of Turkey, responsible for culture and tourism affairs in Turkey. Data on tourism statistics as well as on Cultural Heritage are available through this data source.
3. **Dordrecht**:

- StatLine (see ID *11002* of the DMF, *P_7_UMAS* sheet) is the Statistics Netherlands' database, offering a wealth of data on the Dutch economy and society. From inflation to population, data are clearly presented as they are classified by theme, available to everyone.

4. **Baixa-Limia Serra**:
    - INĒ (Instituto Nacional de Estadística) (see ID *11004* of the DMF, *P_7_UMAS* sheet) is the official agency in Spain that collects statistics about demography, economy, and Spanish society. It is an autonomous organization in Spain responsible for overall coordination of statistical services of the General State Administration in monitoring, control and supervision of technical procedures.
5. **Sava River basin**:
    - For each of the countries taking part in this Open Lab (Slovenia, Croatia, Bosnia and Herzegovina, Serbia and Montenegro) a national database for statistics is available (see *P_7_UMAS* sheet of the DMF).

As more in detail described in section 4.6, data at local scale in general represent a valid source of information, with a level of accuracy which is for sure higher than the one reached through data at global scale. Despite this advantage, local data sources often are characterized by a language barrier, which makes it difficult to consult and select the required data. Anyway, most of the above mentioned data sources have at least some web pages available in English, but this only concerns the national statistical data provider. In the case of the Department of culture websites of certain countries (*e.g.* Bosnia, Turkey, Spain) such an option could not be covered, with all the difficulties this implies.

4.5.2 Classification of resources

The methodology applied within this task foresee a classification of the identified resources according to the following criteria:

- Availability: open/public resources, paid resources
- Relevance of the resources to SHELTER scope
- Identification and analysis of existing systems of population protection of EU countries, including the system of CH protection (integrated rescue system: definition of countries that have it in place)
- Identification of EU countries with dedicated disaster organizations
- Ownership and administration of cultural monuments (state, private), classification, definitions and differences in individual EU countries; existence of a separate central authority
- Price of construction: classification of price categories
- Identification of national budgets of individual EU countries and share of funds for protection, rescue, reconstruction, remediation

- Tourism, culture, social events: how the industry is affected in the event of a disaster
- Demographics
- Expenditure of the state and municipal authorities on remediation and restoration of cultural monuments
- State priorities: health, economy, production, population supply, protection of cultural monuments. Order of importance, characteristics of state interest.

4.5.3 Data evaluation system and methodology

A natural disaster is a rapid natural process of extraordinary proportions, which is responsible for human casualties and great material damage. This process is caused by the effects of gravity, earth's rotation or temperature differences. Disasters affect solid land, water and the atmosphere.

The essence of all-natural disasters are four main processes:

1. rapid mass movements (earthquakes, slope processes)
2. release of deep earth energy and its transfer to the surface (volcanic activity, earthquakes)
3. increase of water level of rivers, lakes, and seas (floods, sea floods, tsunamis, release of glaciers)
4. balancing temperature differences in the atmosphere (hurricanes, tropical cyclones, occlusion fronts).

There are several statistics where experts reflect on the number of human victims and the consequent material damage. *E.g.*, according to UNESCO statistics, every hundred thousand people on earth will lose their lives in a natural disaster. Material damage is not quantified exactly, as no country can quantify the true extent, impacts or damage resulting from natural disasters.

The biggest problem is the statistical provision and subsequent evaluation of historical records of catastrophic events in the studied region, their impact on socio-economic development of societies and the use of socio-economic statistical databases.

It should be emphasized that both in literature and in legal documents of virtually all countries, the terminology covers not only natural disasters and calamities, but also man-made crisis situations from a technical point of view (Chernobyl), a combination of natural and human conditions (Fukushima) and, unfortunately, terrorist attacks (11.9.-USA, Palmyra).

At the same time, virtually all countries declare human lives and health first in their constitutions, followed by the protection of property and the environment from natural disasters.

The following table summarizes the risks directly and indirectly observed in SHELTER with respect to the related indicators' categories:

Directly observed in SHELTER OL	**Data sources, indicators and comment**
Earthquake	Date, Source area, Depth / km, Magnitude, Max. intensity, Origin
Meteorological Storms (Compensation of temperature differences in the atmosphere)	Hydrometeorological data, historical data on temperature observations
Hydrological Floods (River floods, Sea Floods)	Date / Place, a – amount of precipitation, b – duration of rain and floods, consequences, Sea level changes during year
Slope movements (landslides)	Verbal expression of speed of movement, Speed of movement, category I – III.
Climatological Heat waves	Length, Temperature, Thresholds settings, Area, Date, Time
Wildfire	Fire protection brigades data on fires – Dates/ Place, Damages, Occurrence in area
Indirectly observed	**Additional indicators for socio-economic evaluation**
Volcanic activity	Date, Source area, Max. intensity, Origin
Avalanches	Level (1 to 5, according to international classification), Nature and stability of snow cover, Probability of avalanche release
Risks of wind exposure to the earth's surface	Dust and sandstorms
Rapid natural declines of the earth's surface	Karst
Risks of artificial surface subsidence due to mining and mining	Case studies and possible similarity and inspiration from countries – for the purpose of application to OL areas – especially CZ, PL, D
Risks of rock radioactivity	State Office for Nuclear Safety
Radon risk	State Office for Nuclear Safety
Risks of chemical pollution of the environment by natural processes	Mining and processing of metals, industrial production and construction activities
Risk of global warming	
Risk of changes in the Earth's magnetic field	

Table 8 – Risks and respective data categories of indicators

In conclusion, the following monitored historical records of events and socio-economic statistical databases will be analysed and sorted:

- **Earthquake** – Date, Source area, Depth / km, Magnitude, Max. intensity, Origin

- **Volcanic activity** – Date, Source area, Max. intensity, Origin
- **Slope movements (landslides)** – Verbal expression of speed of movement, Speed of movement, category I – III.
- **Avalanches** – Level (1 to 5, according to international classification), Nature and stability of snow cover, Probability of avalanche release
- **Compensation of temperature differences in the atmosphere** – hydrometeorological data, historical data on temperature observations
- **Risks of wind exposure to the earth's surface** – Dust and sandstorms
- **River floods** – Date / Place, a – amount of precipitation, b – duration of rain and floods, consequences, Sea level changes during year
- **Rapid natural declines of the earth's surface** – Karst
- **Risks of artificial surface subsidence due to mining and mining** – case studies and possible similarity and inspiration from countries – for the purpose of application to OL areas – especially CZ, PL, D
- **Risks of rock radioactivity**
- **Radon risk**
- **Risks of chemical pollution of the environment by natural processes –** mining and processing of metals, industrial production and construction activities
- **Risk of global warming**
- **Risk of changes in the Earth's magnetic field**
- **Climatological Heat waves** – Length, Temperature, Thresholds settings, Area, Date, Time
- **Wildfire** – Fire protection brigade's data on fires – Dates/ Place, Damages, Occurrence in area

The proposed methodology applied for the socio-economic data analysis is expressed in the datasets described through the DMF, among which it is possible to find, for example, data sources related to the fire systems and organizations present in the OLs' countries and fire protection brigade's data on fires (see datasets *11020, 11021, 11022, 11023 11027, 11028, 11029* and *11030* of the DMF).

During the next steps of SHELTER project, the proper analytical methods will be identified in order to implement the described methodology.

4.6 Local data sources

According to the GLOCAL paradigm (see D6.1_GLOCAL_V1.0), local data can explain the effect of global climate phenomena that, together with socio-economic information available in the case studies, are worthwhile to give local answers to global problems. This is the main reason why, for each SHELTER case study, local solutions for data and knowledge gathering are evaluated and, in some cases, are going to be implemented: *i)* existing local data network of sensors have been identified and will be integrated in the platforms and tools being developed within SHELTER for the Open Labs, and *ii)* local knowledge from crowdsourcing, solutions based on social networks and chatbots will be implemented.

All the local data sources so far collected and described through the Data Mapping Form derive from a continuous interaction with the relevant stakeholders, through the OLs referent contacts and the so-called '*problem owners*', that are taking advantage of SHELTER activities to share various issues encountered in managing the disaster risk impacting their natural and cultural local areas and to highlight the possible gaps to fill. At the same time, making use of local data sources allows the SHELTER consortium to access a kind of information that is of paramount importance for what concerns historical events. In many cases, local societies have in fact developed their alarm systems, local indicators and traditional recovery systems in different urban and territorial contexts. The collection of such information is also crucial for keeping track of the socio-economic impact that disaster events can have on the local population.

In general, the relevant local data sources identified within Task 1.1 activities have shown to be extremely useful thanks to the high level of information they provide at a local scale, the data robustness and the quite advanced level of maturity. This is for example the case of the Sava River basin OL, for which a common platform (the Sava GIS) for sharing and disseminating information and knowledge on water resources protection and water management activities in the Sava River basin is available. Sava GIS is composed of contributing geographic information systems of ISRBC member countries (Slovenia, Croatia, Bosnia and Herzegovina and Serbia) and is representative of many aspects that concern the local data retrieval:

- **Local data standards**: to make data interoperable (which is one of the main principles in SHELTER data management) the datasets should be available in a standard data format and exposed by a standard data access service. In fact, data can show many different formats implementing different information models. Within Sava geodatabase, hydrological and meteorological real-time and processed data has been designed and structured in accordance with OGC WaterML 2.0 standard which is used for the representation of water observations data, with the intent of allowing the exchange of such datasets across information systems. This is a central point when it comes to making the SHELTER Data Driven Platform communicate with other already existing data sources and not all the local providers expose such an advanced level of development in terms of data standards.
- **Local data access**: to prevent data access by unauthorized and unknown users it is necessary to foresee a proper data access management to guarantee data security for what concerns both sensitive and non-sensitive data, as well as for out-of-scope data use. In the case of Sava GIS, ISRBC don't represent the data provider nor the data owner, although they have the role of data exchange coordinator on the Sava River basin level, providing the data exchange platform

for that purpose. Therefore, allowing the SHELTER consortium to access the ISRBC systems is a very sensitive issue and requires a specific procedure. SHELTER consortium has been allowed to access the Sava GIS data only for the purpose to develop compatibility of the ISRBC systems with the SHELTER DDP. The access and visualization of data have required an official approval of the Sava GIS contributing countries and the application of a specific procedure, through the signing of a user declaration and the consent to use the related data for purposes of analysis and studies within the SHELTER project, taking into account the specific limitations and/or conditions for implementation and exploitation of the ISRBC data (from Grant and Consortium Agreement). Through the described procedure a certain number of ISRBC WebGIS Viewer accounts (*i.e.* user with rights to view and download data) has been assigned to those SHELTER partners that needed for example to retrieve information on historic events (POLITO) or on data available from IoT sources (RED and POLITO).

- **Local data maturity**: the example of Sava OL is quite peculiar as it shows an advanced level of maturity, particularly in terms of data standards and data availability. Anyway, not all the 5 SHELTER OLs show the same level of maturity, as already anticipated in the introduction of the present document.
- **Local data reliability**: among the functionalities provided through the Sava GIS platform a validation tool of the data uploaded to the system is available, allowing the user to check on the reference system for example and to transform it if necessary. This of course makes the data more reliable, as it ensures that the performed geographic transformation is correct and in compliance with the applied standards. Anyway, there are many aspects that concern local data reliability, and this is also highly dependent on the type of data at hand. Among the local data, social media posts for example are unstructured and extremely volatile in terms of accessibility. They can suffer from disinformation, false judgments, social biases thus making the produced information deeply unreliable sometimes. On the other hand, social media data has the capability to bring the local knowledge from the territory and citizens to decision makers, providing local reports and insights in data-poor regions. Citizen science reporting via social media and other platforms can radically expand scientists' observations of ecological systems.
- **Local data technologies**: among the various aspects of local data, the technological maturity makes the five OLs differ from each other significantly. The main gap lies on the presence of WMS and/or on the exposure of APIs. Local data provider such as the abovementioned '**Sava GIS**', the 'Información Xeográfica de GALICIA' (Baixa-Limia Serra OL) and the 'Nationaal Georegister – NGR' (Dordrecht OL) give the users the possibility to retrieve the available data through WMS services, thus allowing SHELTER technology partners to quickly import geospatial data in the tools and models they are developing within the project. Unfortunately, not all the SHELTER OLs are characterized by such an advanced level of technologies for data sharing. As anticipated in section 3.1.1.1, the availability of

this kind of services represented one of the main selection criteria applied to the DMF.

- **Language of the local data source**: Information retrieval deals with finding useful information from a large collection of unstructured, structured, and semi-structured data. In the current scenario, the diversity of information and language barriers are the serious issues for communication and cultural exchange across the world. Information retrieval can be classified into different classes such as monolingual, cross-language and multilingual information retrieval (Dwivedi, S.K. and Ganesh, C., 2016 [7]). Among the local data sources available from the five SHELTER OLs, again the Sava River basin shows an advanced level of maturity for what concerns the language aspect. Despite it represents a case of monolingual information retrieval, the used language is the English one, which is becoming more and more the language commonly utilized for communication among countries. For the Baixa-Limia Serra, the Dordrecht and the Ravenna OLs, on the contrary, the identified local data sources are characterized by the exclusive use of local languages (Spanish, Dutch and Italian respectively) for data information and instructions on how to retrieve them, thus making hard to perform the selection of data considered relevant for SHELTER. This issue has been tackled by directly involving some of the OLs referent contacts in the action of filling the DMF with datasets descriptions in English language. This methodology has been helpful in the data selection phase, but the language barrier still persists when the technology partners need for example to access the data and the related documentation (*e.g.* in the case of IoT data).
- **Local data integration**: as already anticipated in sections 4.2 and 4.3, local data should be combined with data at global scale, as long as they are compatible. Certain local data could also require integration with further information on specific topics, which is the case of the Sava GIS data. As more thoroughly described in D1.2 (see section 6.7.1), in fact, despite the Sava River basin OL was provided with many available DRM data already before SHELTER was launched, the cultural-historical heritage located in this OL region was still missing a geospatial format. This is the reason for designing the previously mentioned mock-up for a CH attributes template in the framework of Task 1.2 that will be further developed in Task 2.3. With this regard, the **Ravenna** case study, because of its Italian background and its strong focus on the cultural heritage domain, turns out to be very complete in terms of both cultural and climate data availability. The Regional Agency for Prevention, Environment and Energy of Emilia-Romagna (ARPAE) has the aim of controlling the state of the environment and supporting the sustainability of human activities, aiming at the protection of human health and territorial competitiveness. It then represents a valid data source at regional scale, particularly for what concerns meteorological historic and forecast data, air quality information and cartographic data. Moreover the Italian cultural heritage domain looks pretty advanced in terms of digital innovation, thanks to the ‘Vincoli in Rete’

programme, which was released by the 'Istituto Superiore per la Conservazione ed il Restauro' and the 'Istituto Superiore per la Conservazione ed il Restauro' (MiBACT).

In conclusion, the local data sources so far identified for the five SHELTER OLs show, together with the already described advantages, also some criticalities. These are mainly represented by a limited data accessibility, due to the lack of standardized platforms and the related data, language barriers as well as a different level of technological maturity and of data validation (data reliability). On top of this, each OL seems to have a different degree of awareness of how crucial it is to put effort in protecting both the Cultural and the Natural Heritage present in the respective countries, or in properly balancing such effort between the two types of heritage. As also underlined in the '*Copernicus services in support to Cultural Heritage*' report, in certain cases site management lacks proper institutionalisation of the use of site management plans, which weakens site monitoring, economic support and consequently data collection. It appears that this phenomenon can be differentiated if it is Tangible Heritage or Natural Heritage that is considered. In fact, even if data is available for European Natural Heritage sites, it seems that site operators working on Tangible Heritage lack data and thus they cannot implement or share good practices. As a consequence, answering to the demand of Tangible Heritage site operators could appear as a priority. Even though collected data is easily used for monitoring and site management purposes, the Cultural Heritage communities intervening in Tangible Heritage are lacking a clear process of site monitoring. By comparing needs in both environments and identifying their similarities, one key conclusion is the coherence between this more integrated approach in which Tangible Heritage could benefit from best practices from the Natural Heritage community, in order to foster its global development and sustainability.

Building long-term resilience takes more than enhancing the ability of both external and local actors to react to single events. Resilient communities manage their natural systems, strengthen their infrastructure, and maintain the social ties and networks that make communities strong. The use of local data, particularly social media data, has the advantage of raising citizens' awareness and empowering them to take action, thus increasing communities' engagement. In many countries there are well-established systems and tools used for inventory and documentation of cultural heritage. They reflect the tradition of cultural heritage protection and the local approaches and understandings condition their content structure. Nevertheless, in some countries there are several systems for data collection, which are not connected together. Therefore, the straight comparison of data on heritage assets is not possible. The responsibility for collecting data depends on the administrative structure in each country. Planning of broad activities, such as preventive strengthening or even post-earthquake measures in earthquake prone areas, or energy preservation measures, can be better based on mutually developed methodology (Zarnić *et al.*, 2017 [8]).

Being the five SHELTER OLs so different from each other under many of the analysed aspects, the main challenge of this project in terms of data sources and knowledge collection lies in developing a homogeneous system starting from a quite heterogeneous context. Table 9 shows the level of maturity of each OL with respect to the previously analysed data aspects:

Open Lab	Standards	Access	Maturity	Reliability	Technologies	Language
Ravenna	green	yellow	green	green	yellow	red
Dordrecht	green	green	yellow	green	green	red
Baixa-Limia Serra	green	green	yellow	green	green	red
Sava River basin	green	yellow	green	green	green	green
Seferihisar	grey	grey	grey	grey	grey	grey

Table 9 – Level of maturity of the SHELTER OLs (with the exception of Seferihisar) with regard to the main local data aspects (*green=high, yellow=medium, red=low, grey=to be verified*)

Despite the big effort demonstrated by the Seferihisar OL referent contact, interactions with the local stakeholders have been challenging, due to a slow response on one side, and to a change in the local referent contact after some months of interactions on the other side. The situation has been worsened by the COVID-19 situation. For these reasons, it was not possible to understand whether local data sources are available for this OL (with the exception of few statistical data sources identified by UMAS). This represents a gap which needs to be filled in the next steps of the project. Anyway, it was still possible to identify several data sources at global or European scale (*e.g.* the UERRA dataset – ID *73005* of the DMF), which can temporarily compensate for the lack of local data sources (see Figure 18).

Figure 14 summarizes the distribution of the local data sources so far identified for SHELTER OLs (with the exception of Seferihisar):

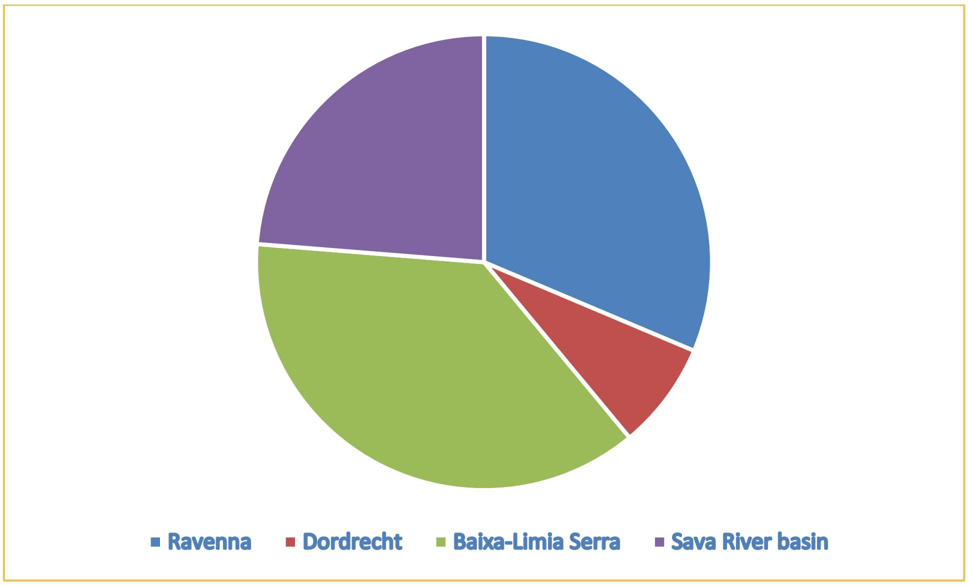


Figure 14 – Open Labs local data sources distribution (with the exception of Seferihisar OL) as extracted from the DMF

Figure 15 shows what the current availability of local data sources is, in comparison with general purpose data (*e.g.* Copernicus data, EM-DAT data, etc.).

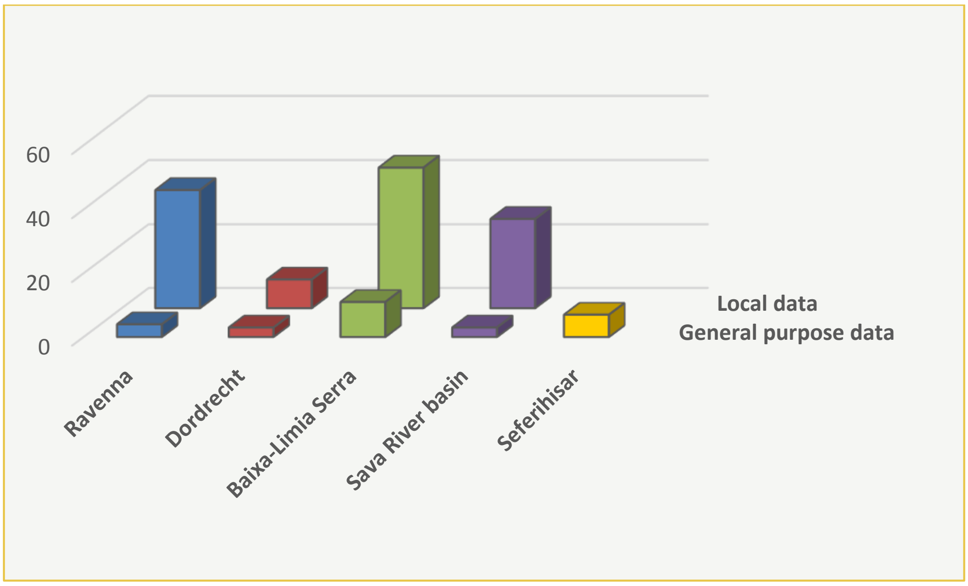


Figure 15 – Local data sources versus general purpose data distribution among the five SHELTER OLs as extracted from the DMF

Apart from the case of Seferihisar OL, the majority of the data so far described in the DMF derives from local sources. This represents a strength for the level of details that such data can provide. At the same time, it has to be taken into account that local systems generally show a certain resistance to change, as they have been designed for a specific purpose and, in this sense, they can be not very flexible. Non-local systems are designed for general purposes and can more easily adapt in case of environmental changes for example. For this reason, local data technologies should be designed with a mid/long-term perspective and foresee possible modifications in line with the new situations, in order to make the overall process of adaptation easier.

5 SHELTER knowledge for CH climate and disaster resilience

In the framework of WP1, Task 1.2 **– Codification of existing knowledge** has the objective to extract, structure, share, operationalise and take advantage of all the existing knowledge, including local knowledge and social memory, best and next practices, and linked research initiatives. The task includes the identification of existing specific ontologies and controlled vocabulary dealing with the domains of CCA, DRM and CHM and tailored to the project requirements at the same time. The protocol defined is used by the OLs to collect relevant information.

As well as the data identified within Task 1.1, all this information has been identified, described, filtered and assessed with respect to the requirements, the indicators and the expertise of WP1 involved partners (see Figure 3). Initially a common data gathering template has been structured to facilitate the collection of existing knowledge; following the same methodology illustrated in section 3, such knowledge has been then described by means of four dedicated structured data description tables, with cross-cutting characteristics included in all the four templates. Such tables successively converged in the Data Mapping Form as datasets produced by the related technology partner, according to the following schema (Table 10 – Mapping of the different outputs from Task 1.2 onto the DMF):

DMF sheet name	DMF dataset ID	Dataset description
P_2_UNIBO	12001	Best and next practices
P_3_UNESCO		Regulatory framework
P_4_POLITO		Historical events and social memories
P_10_CRCM		Linked research initiatives

Table 10 – Mapping of the different outputs from Task 1.2 onto the DMF

All the outputs produced within Task 1.2 will be ingested in the Data Lake, but only the '*Historical events and social memories*' output will be integrated in the Resilience Dashboard, as it includes geospatial data (*i.e.* geolocalised disaster events). The attributes collected for each past disaster event in fact will come together with the coordinates of the centre of the impacted area. In this way it will be possible to locate such historical events on a dedicated map that will be linked to the Resilience Dashboard.

In the following sections the main concepts and principles applied to each of the above mentioned types of collected knowledge are summarised, with particular focus on the role played in the framework of Task 1.1. Refer to D1.2 – **Building of best/next practices observatory** for a more comprehensive description.

5.1 Ontology

The main scope of the ontology is to solve the interpretative gaps that can arise when shifting from a virtual reality to a concrete one. In general, the data themselves are not sufficient to properly describe and analyse an existing problem and this is even more challenging when it comes to cope with such a complex task, like the one representing the main SHELTER objective: the resilience and sustainable reconstruction of historic areas, taking into account climate change and hazard events.

An ontology formally represents knowledge as a hierarchy of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts and allowing the connection between different domains.

In this sense, ontologies are the structural frameworks for organizing information and are used in artificial intelligence, the Semantic Web, systems engineering, software engineering, biomedical informatics, library science, enterprise bookmarking, and information architecture as a form of knowledge representation about the world or some part of it. The creation of domain ontologies is also fundamental to the definition and use of an enterprise architecture framework (Rogushina, J. *et al.***Errore. L'origine riferimento non è stata trovata.**, 2018).

Figure 16 describes the role of the ontology in the data analysis life cycle:

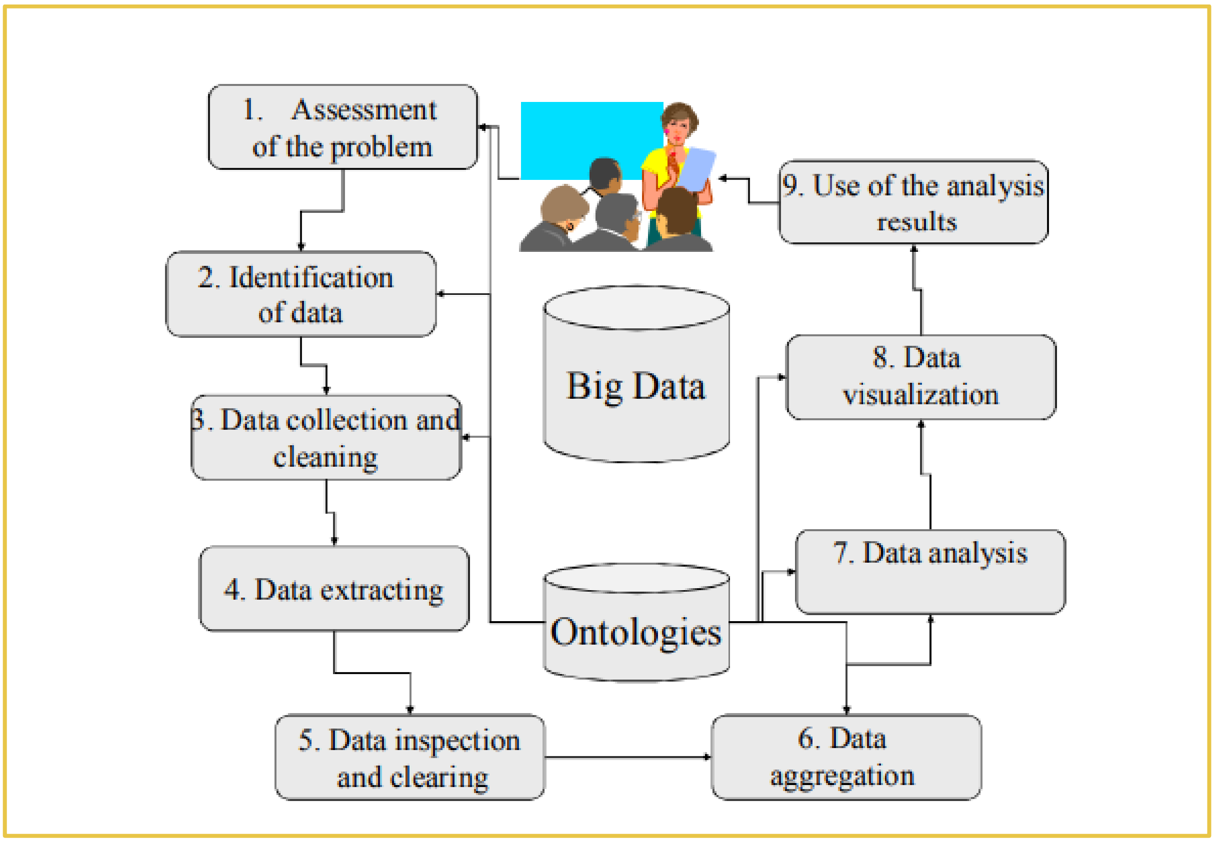


Figure 16 – Role of the ontology in the data analysis life cycle

At the stage of **data identification**, the datasets necessary for carrying out analytical projects (tasks) and their sources are defined. Domain ontology helps in identifying appropriate data sources.

At the stage of **data collection and cleaning**, the final formation of Big Data packages for the purposes of the task is accomplished using semantic analysis of metadata text annotations and the selection of relevant datasets for solving the problem. Semantic approach is thus used for selection of Big Data sets that are relevant to the user's task.

The stage of **data aggregation and presentation** serves for consolidation of datasets that can be distributed across multiple datasets through common fields, such as by dates or identifiers. In other cases, the same data fields can be displayed in multiple datasets. In any case, completion of this stage can be complicated due to differences in: *i)* data structure, *i.e.* although the data formats may be the same, the data structure model may differ; *ii)* semantics, *i.e.* different values in two different sets of data may mean the same thing. At this stage domain ontology can be used for matching of various names of the same concepts, for determining of relations between them (hierarchy, synonymy, semantic closeness, etc.).

It is clear then how ontologies are of fundamental importance when exploiting Big Data, which are largely used in SHELTER. Both the task definition and the annotations of Big

Data are natural language unstructured or semi-structured texts. Therefore, their matching can be based on methods of natural language analysis, but through the Big Data ontology, which contains knowledge about the specifics of this domain and allows semantic processing of other elements of Big Data metadata (to match the parameters of the metadata structure with the domain concepts).

In order to identify relevant ontologies for SHELTER, the review of the existing literature has been applied as the main methodology (see section 4.2 of D1.2). The main criteria to determine relevance was the closeness to three domains: Climate Change Adaptation, Disaster Recovery Management and Cultural Heritage Management. The results of this process are the SHELTER ontology itself and the SHELTER core vocabulary, consistent with the ontology, but simpler to use when the ontology is not required.

Every term in the SHELTER ontology is part of the SHELTER core vocabulary. These terms are defined in Table 3, section 4.3.1 of D1.2. The SHELTER core vocabulary, which is simpler to use than the ontology when this is not required, has been also included in the SHELTER Wiki page [AD.2], conceived as a tool to be used not only by SHELTER Open Labs, but available for the general public. Each vocabulary has been listed with its own definition, and tags (*e.g.* types of hazard addressed, relevance for the SHELTER Open Labs) have been added to better filter the vocabulary contents.

The ontology developed in SHELTER will be continuously cross-checked, validated and complemented by the Open Labs.

For a more comprehensive description of the applied methodology and the results obtained through the definition of the ontology, refer to section 4 of D1.2.

5.2 Regulatory frameworks for resilient cultural heritage

In the context of Task 1.1, the importance of the analysis and definition of the regulatory frameworks for resilient cultural heritage lies in the opportunity of applying specific existing standards during the collection of good practices. This way it will be possible to start from already defined policies, emergency protocols and strategies adapting them to the five SHELTER OLs (*i.e.* developing a global database focused on case study countries) instead of designing new regulatory frameworks from scratch.

This approach was applied also during the design process of the mock-up for CH attributes template requested by the Sava River basin OL (see section 6.7.1 of D1.2), for which the possibility to base the mock-up drafting on already existing definitions provided by UNESCO and ICOMOS (International Council on Monuments and Sites) for example made the whole final result more robust.

In the framework of Task 1.2, UNESCO undertook a comprehensive review and codification of existing knowledge in DRM for CH. The objective was to build a global regulatory framework database as a common reference for defining replication conditions. The exercise consisted not only in mapping, but also in analysing the last 20 years of DRM frameworks.

The performed desktop analysis, though not exhaustive, provides a relevant overview on the available documents that address DRM for the CH sector (Disaster Risk Reduction, emergency preparedness and response and post-disaster reconstruction phases).

The results have been presented through graphical representations (see section 5.3 of D1.2), organizing them per document's type, document's scale, hazard addressed by the document, object scale (document's quadrant distribution) and resilience's scope distribution. Moreover, the reviewed documents have been divided into two main groups, depending on the scale: at international level and at EU level.

Also used keywords have been analysed (and this connects to the importance of the abovementioned concept of ontology), highlighting how '*world heritage*' and '*climate change*' have been the most used.

As already emerged during the SHELTER workshop on '*GLOCAL user requirements for Disaster Risk Reduction (DRR) and Heritage*' that was held in Venice last year, in many countries there is a lack of adequate risk assessment procedures and most of the international-level and European-level documents that were reviewed only refer the importance of adopting adequate risk management practices for CH protection, highlighting also stakeholders that should be involved, but mainly representing a sort of recommendation, with almost no guidelines.

Only some of the countries that were analysed, in fact, have developed guidelines and legislation that can be applied to support the implementation of DRR practices for certain hazards. Few documents address practical frameworks, methodological approaches or more detailed guidance for implementing risk management for CH. For this reason, there is room for developing better risk assessment and risk mapping procedures in the future, for CH addressing different hazard types.

In terms of post-disaster recovery/reconstruction of CH, most of the relevant documents covering this issue were produced at the international level.

The potential of using the OL workshops within the SHELTER project framework as a platform for discussing about those findings would represent a perfect basis for sharing knowledge and experiences. This analysis is not exhaustive, considering many National documents are not in English and precious would be the help of national stakeholders. This aspect again reflects the disadvantage that derives from the use of local data and knowledge (see section 4.6), which implies dealing with language barriers most of the time, despite the advantage of accessing more locally tailored information.

For a more comprehensive description of the applied methodology and the results obtained through the review of the available regulatory frameworks, refer to section 5 of D1.2.

5.3 A collection of best/next practices and tools

SHELTER Task 1.2 is fully dedicated to the codification of the existing knowledge, including relevant best practices (practices already validated and demonstrated), next practices (innovative practices not totally documented, but promising and inspiring), as well as tools and methods from linked R&I (Research & Innovation).

The outcome of this work will be an observatory with the scope to support Open Labs in the implementation of their project activities.

The following summary is meant to provide a picture on the results of the performed survey with a view to Task 1.1 activities. Refer to section 6 of D1.2 for a more comprehensive description of the methodology applied and the general conclusions.

For what concerns the **best practices**, the result of the research performed by UNIBO is a collection of 67 between good and next practices coming from related EU projects, among which 60 of medium or high relevance for SHELTER. The outcome of the observatory can be consulted in Annex II of D1.2. From the analysis of such results, it is clear how, over the last 20 years, the importance of involving stakeholders and citizens in the DRR process has grown in awareness. Stakeholders represent a fundamental contribution to the DRM actions as direct actors dealing with the HA.

Among the practices collected, a very broad group includes the use of the new technologies to be used in several phases of the DRM and many have stressed the value of creating digital simulation models.

For identification and analysis of international projects in the sense of **tools** applied in EU projects and **linked R&I initiatives**, a specific template was developed by CRCM in close coordination with UNIBO. The aim was to structure the available information via using several databases and project descriptions in order to create a wider overview of the existing tools.

As a preliminary result of the survey on tool type, among the more successful or more promising R&I initiatives a combination of modelling system and forecasting/monitoring/(early)warning/control/decision support system seem to be more dominant. To a larger extent, these systems or combinations of systems are based on GIS and/or satellite services. Solutions regarding social media as well as mobile applications are seemingly not so often the objectives of these projects and Artificial Intelligence (AI) and Virtual Reality/Augmented Reality (VR/AR) are still scarce.

For what concerns SHELTER multiscale character (according to which artefacts, buildings and archaeological sites have been grouped and named as the object/building scale, while neighbourhoods/districts, cities and regions have been assigned to the urban/territorial scale), the majority of good practices and tools' scale documents refer to multiple scales. All the scales are almost equally represented and none of them prevail on the others.

When analysing the observatory under the point of view of hazard related practices, the majority of results are addressed to floods, a hazard currently deeply studied and with a great variety of solutions and methodologies available validated in pilot cases (see Figure 28 of D1.2).

With this regard, the main gaps identified are related to two relevant hazards for the SHELTER project: subsidence and heat waves. For both these two kinds of hazards the reason could lie in the close relationship they have with climate change, which is reflected in the global mean sea level and in the global warming. Particularly the heat waves can be considered as a quite new phenomenon and only recently has caught the attention of meteorologists and climatologists.

The results derived from these surveys (best practices and R&I initiatives) are strongly related to the data collection monitoring in SHELTER. Figure 17 shows the distribution of the datasets identified so far (and described through the DMF) for the five OLs in terms of their relevance for the hazards investigated in the project:

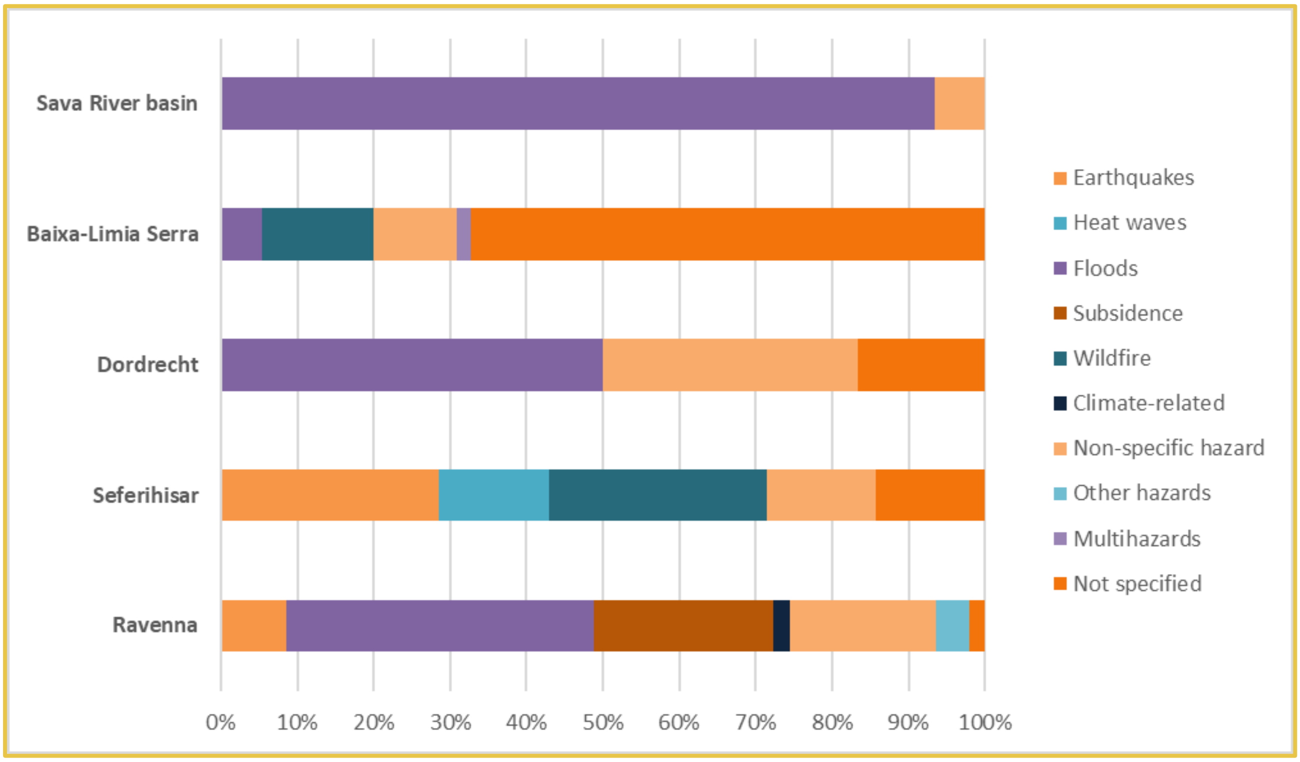


Figure 17 – Data distribution of the SHELTER OLs in terms of the hazards investigated in SHELTER

Such distribution of course reflects the role that each hazard plays towards each OL (see Table 1), when looking at the graph as a whole, which is: being for example the Sava River basin OL affected by the flood hazard, it is expected to find a high number of flood-related data described for this OL. Anyway, looking at the availability of data within each OL, the proportion of subsidence- and heat waves-related data versus the data available for the other hazards is clearly unbalanced, mainly towards the flood hazard, thus showing a coherence with what emerged from the survey performed on existing best practices and R&I initiatives.

In conclusion, it is expected that, when looking for further best practices related to other hazards than subsidence and heat waves, with high probability one will find many sources, which most likely implies the same availability in terms of related data. This means that all the solutions that will be implemented for the more investigated hazards will probably lack of innovative approaches, since a lot has been already done in other practices.

On the other hand, the new best practices that will be developed in the framework of SHELTER for those OLs impacted by the so far less analysed hazards will be necessarily characterized by a pioneering approach.

5.4 Historical events and social memories in SHELTER Open Labs

A resilient community learns from past disasters about its own vulnerabilities. It develops the capacity to detect and monitor emerging hazards and vulnerabilities that may build up over time. It then acts to reduce its vulnerability by improving the management of natural systems, strengthening infrastructure and social networks.

As underlined in D6.1_GLOCAL_V1.0, local and traditional knowledge about the experience (good or bad) gathered from recent disasters in certain countries is unfortunately scarce. Despite the high value of historical events memory, this kind of knowledge remains not easy to gather because it includes a wide range of heterogeneous information with various grades of impacts and reliability. Difficulties are mostly due to the identification of a common coherent system to measure and provide precise information at different periods of the past and in different countries.

In this aim POLITO has defined a protocol to collect information about historical catastrophes and risks into a template to be shared by considering a temporal framework. The protocol defined will be used by the Open Labs to collect relevant historical information about the SHELTER use cases, allowing the retrieval of data according to common indicators of description and measurements.

The need for a shared framework, language and system of identification is especially important referring to past events. The approaches to disasters have been culturally and scientifically different through time and space.

The methodology developed by POLITO includes desk research, a questionnaire for OLs, and the definition of a template as protocol for collecting heterogeneous data on risks and disasters. By this approach an accurate bibliographical research and data analysis has been developed.

As first step, a definition of a time framework for the 'past' was needed. Data collected from the databases mostly dates to the 1980s.

The desk research has consisted in listing existing data sources such as scientific literature and existing databases in different countries that have recorded those past events. The Protocol provides the link to the databases.

Currently, there are a limited number of national and international databases that provide information about historical events; however, most of these databases limit their time frame to recent history. The most widely used international historic events database is EM-DAT (Emergency Events Database) (see D1.2 for further databases; see also ID *72005/73002/74009/75046/76006* of the DMF for a detailed dataset description).

Interoperability of the existing databases with the SHELTER Historic Events protocol is at the core of the methodology, since existing databases already provide some essential information regarding past disaster events that occurred in the sites of each Open Labs. This approach will be moved to a further step with the Task 2.3 – **Anatomy of Historic Areas: collective characterisation of CH assets**, in which the historic events information for linking cultural heritage assets to their intangible (narratives and visual sources) will be included.

The SHELTER protocol to collect historical events and social memories is based on a comprehensive attributes table organized through an excel workbook. The attributes of the template have been defined in a way that the Open Labs could be requested to provide also historical visual, audial, written documents in addition to compiling the protocol. The communication process with the OLs has been triggered and mediated by the WP1 leader, through a Level 3 interaction procedure, according to which the Open Lab Task Working Group (OLTWG) is asked to interact with the targeted stakeholders forwarding the requests raised by the technical partners.

At the time of writing the present document historical events and social memories information was available for Ravenna, Dordrecht, Baixa-Limia Serra and Sava River basin OLs. The next step is to control and to study the above mentioned local, regional and international databases to operationalize the protocol. This way, it will be possible to create a coherent structure regarding the diverse nature of each Open Lab in terms of scale, hazards, management, etc.

A detected gap is related to the amount of visual, written, audial documentation with respect to the social memory. In fact, the amount of information obtained from the Open

Labs is not sufficient to generate a coherent and well-balanced narrative regarding each Open Lab.

As highlighted in the '*Copernicus services in support to Cultural Heritage*' report, digitisation and digital preservation through the development of 3D-capturing, 3D-processing and tools for text digitisation or preservation of audio-visual material is being in general developed, but it is estimated that only around 20% of Europe's collections have been digitised so far, leaving therefore about 80% of resources still to be digitised. In addition, local knowledge could change with time and vary from regions. It therefore needs to be developed and improved so that it could be applicable and sustainable, *e.g.* traditional techniques. It looks then crucial to digitalize the information and make it available for the public, also because digitisation and online accessibility lend Cultural Heritage a much greater visibility.

The collective memory/experience notably for historic events/disasters should draw the attention of the policymakers and practitioners and be integrated into local practices.

6 Data replicability

Data replicability is a relevant concept in data management as it allows the technological progress, the advancements in sharing innovative solutions beyond the domain (thematic, geographical, scientific, technological, ...) and the scope for which the results have been achieved. Moreover, it increases the impact of the achieved results producing data that are replicable in other contexts.

The CH sector has traditionally been concerned with sharing resources and furthering human knowledge, with particular interest to the issues associated with metadata and interoperability (which is strictly related to data replicability), especially when it comes to the use of technology. These goals and interests in the CH sector are natural alignments with those of linked data; hence, there has been an increasing interest in the application of linked data in this sector. Anyway, although the concepts of linked data and semantic technologies have been around for some time, its adoption is still at an early stage, and sufficient maturity of these technologies for widespread application and adoption in sectors such as CH is questionable. There remains work to be done in assessing the needs of this sector in relation to these concepts, the success of projects already implemented, and potentially significant further work to be done to rectify any gaps which prevent the sector from using the technology to its full potential (Davis E. and Heravi B., 2021 [9])

6.1 Definition

Ostermann and Granell (2019), in their '*Reproducibility and replicability in Science*', make a useful distinction between reproducibility and replicability:

Reproducibility *is … concerned with the validity of the results of that particular study, i.e. the possibility of readers to check whether results have been manipulated, by reproducing exactly the same study using the same data and methods.* ***Replicability*** *is more concerned with the overall advancement of our body of knowledge, as it enables other researchers to conduct an independent study with different data and similar but not identical methods yet arriving at results that confirm the original study's hypothesis. This would be strong evidence that the original study has uncovered a general principle through inductive research, which now another study has proven in deductive research design.*

Therefore, reproducibility requires full access to both data and methods used. Replicability is more modest, but not less useful, and requires access to a description of the method or pseudo-code and access to metadata describing how the data was collected and its context, even if the original dataset is not accessible.

In short, reproducibility involves the original data and code; replicability involves new data collection and similar methods used by previous studies.

In SHELTER replicability is privileged, but thanks to the implementation of the Data Management Plan and the application of the FAIR principles to the data identification, data collection, data production and data curation, the reproducibility will be made possible as well.

Replicability is affected by the quality and the attention paid to the design and methodology of the process including data handling. In SHELTER the data identification has been designed and implemented through a process that has involved standards, protocols, partners and stakeholders.

Since SHELTER includes five Open Labs which have both unique characteristics and common problems, the identified data must support the implementation of common tools and methodologies that provide customized solutions, which is the essential of the replicability concept.

6.2 Principles

The principles for a data replicability strategy implementation are partially borrowed from the data management plan and extended with some considerations and good practices.

The replicability is one of the criteria to be considered during the design process of the data collection methodology. It is not easy to assess a dataset as replicable but following the abovementioned definition it is possible to identify three particular characteristics that can suggest how to estimate the replicability: the representativeness, the reliability and the technical aspects.

The **representativeness** deals with the content of the data: if a given dataset has been used to suggest new useful insights and to represent a diffused and large scale phenomenon in terms of thematic domain extent, geographical extent and temporal duration, it is clear how such dataset and the related methodologies or developed applications can be considered more replicable with respect to a dataset representing a local, specific or rare phenomena. The completeness of the dataset is another characteristic to be considered in the evaluation of the representativeness and hence of the replicability. Indeed, an incomplete dataset might bring limited, partial and misleading results.

The meaning of the **reliability** concept is quite intuitive: if a dataset is not reliable cannot be considered replicable. On the contrary, an unreliable dataset can represent a source for errors and for the propagation of mistakes and misinterpretations. Anyway, if on one hand the representativeness depends on the nature of the dataset and of the represented phenomenon, on the other hand the reliability depends more on the methodologies and on the quality of the data generation and collection. Consolidated methodologies, application of standards approaches and transparency in the processes are the principles to ensure a decent degree of reliability in data collection and generation.

The **functional aspects** of replicability deal with the data accessibility both legally and technically, extending to methodologies and obtained results.

Legal aspects include data licences. It is clear that for existing data the license is given and can be rarely changed. Of course, datasets exposed with open licenses, like for example creative commons, could be privileged in the data identification and selection process. For what concerns new data, open access licenses must be strongly suggested to the data owners/producers.

Technical aspects refer to data accessibility or, in other words, to the application of the FAIR principles. Making the data available in machine readable format is a matter of the application of standards for both data and metadata and the application of appropriate ontology to make the data discoverable on the web.

The combination of datasets correctly described and formatted with standards, with open data license associated with scientific and technical publications released in open access is the best way to ensure the replicability of the data.

Long-term perspectives

The long-term perspectives deal with the handling of the data both during and after the end of the project. Whenever a particular dataset is identified, collected and generated it is fundamental to look at the scope of this dataset not only with respect to the specific scope of the project, but also to its potential use beyond the end of the project and in more general terms. Indeed, the dataset could become relevant in ten years as reference data for comparison and temporal trend analysis. Or it could be useful as a complementary dataset for different scopes with respect to the specific project purposes. Thinking the whole life cycle of the dataset in long-term perspective from the generation, the collection, up to the curation phase by applying for example the standards in data format, in metadata and associating an ontology, means to ensure the data replicability.

6.3 Limits of data replicability

Not all the data can be considered replicable. Based on the principles mentioned above it is quite easy to understand the reasons why a certain dataset can be considered not replicable.

The **limited representativeness** is one of the main reasons. Indeed, if the data are describing complex systems or there is a significant level of noise/bias or simply a mismatch between the scale of the related phenomena and the scale at which such phenomena can be measured, the data cannot be considered replicable.

The **scarce reliability** of the data due to lack of quality on data collection or data generation methodology is another cause that can compromise the replicability of the dataset.

The **technical barriers** to find, access and share the data represent a further factor that can make the data not easily replicable. As described before, if the datasets are not duly formatted, exposed and provided with standard metadata, it might be hard to be reused, analysed and replicated.

Among the limits described above, there is a set of further obstacles to be considered which deal with **ethical, political and commercial** aspects. It might happen that the datasets collected and/or generated are sensitive data related to socio-economic issues that concern privacy at individual or community level. In that case the access and the replication of datasets and methodologies might be restricted and subject of Non-Disclosure Agreement (NDA). The sensitivity of the data could be related to political issues in the case, for example, of dataset concerning national strategic sectors, immigration, illegal activities and conflicts.

The commercial aspects are clearly related to the Intellectual Property Rights (IPR), the protection, the copyrights, the licence, the cost of data production, the collection and the curation, the profit in case the data are provided by a private owner. Commercial data are protected by IPRs and License: this means that the data replicability might be limited by access and reproduction costs. To avoid or to mitigate this issue it is fundamental to identify and to define the terms and conditions associated with the datasets.

6.4 Data replicability approach implemented in SHELTER

The data replicability approach of SHELTER reflects the application of the principles above described.

The **representativeness** of the data has been ensured by involving all the technical partners and the stakeholders in the existing dataset identification process with the scope to identify those datasets that can be available for the project. Starting from the preliminary available dataset list, the assessment of the relevance and the usefulness of the identified datasets has been conducted by comparing the data with user requirements and indicators. The degree of connection of the datasets with the requirements and the indicators can give a measure of the representativeness, and hence of the replicability, of the datasets. The replicability of certain datasets could be reduced, like local data and IoT data that are related to sensors locally installed and to the retrieval of local information. Collected data might result in local specific problems so that they cannot be used in other contexts. However, if even those data are made available together with analysis methods and results through open access and open publication, a local dataset can be used as reference data and can inspire the implementation of other similar technological solutions.

The **reliability** of the existing datasets has been ensured by paying attention to the data sources and the data providers. As aforementioned, the reliability of data is a matter of quality of the processes and of standards applied to the data collection and data generation. Data sources documenting the standards and the processes must be privileged in the data selection phase. The majority of the data source providing existing dataset are institutional and public entities that can ensure a decent level of quality in the processes. Other datasets, like those coming from local sources, crowd sources, IoT sensors have to be handled carefully because of their unverifiable acquisition processes.

For what concerns the **functional aspects** related to the data replicability, as above described, the Data Mapping Form (see section 3.1) has been defined as a sort of data brokering tool built on the basis of the SHELTER Data Management Plan with the scope to manage the data during the cycle of the project. The structure and the fields of the DMF to describe the datasets include those parameters that allow to assess the replicability of the data.

As an expression of the Data Management Plan, the DMF reflects all the concepts concerning the accessibility of the data. Indeed, the DMF reports metadata, licences, machine-readable links, tags.

The process to compile the DMF has involved all the partners and the stakeholders of SHELTER by asking them to provide data description in the structured data framework following the compiling instructions. In the first interactions of the DMF population, a technical support has been provided especially to the Open Labs to compile correctly the records, in this way there has been a knowledge transfer and a rise of awareness about the importance of a structured and organized data management approach that will be taken into account for the generation of new data foreseen in the technical tasks of the project.

The relevant information has been identified taking into consideration the complementarity and the redundancy of some kind of data to ensure the applicability of the methodologies implemented in SHELTER. This is for example the case where local data are missing and the same kind of data on the same location are available even though with a lower ground resolution or a lower accuracy. In this way even the replicability is ensured allowing the implementation of similar analysis methodologies with the same kind of data beyond SHELTER.

For what concerns the **long-term perspective**, besides the application of FAIR principles reported into the DMP, the identified data that, in the iterative process described in section 3.1, are assessed as not relevant for SHELTER are not removed from the DMF. They will not be considered as relevant for the specific scope of the project at the stage of the assessment but are kept recorded and classified as '*not relevant*' since they could become useful later on along the project or beyond the project or for other similar initiatives.

7 Gaps and roadmap

The complexity of SHELTER is due to the great number of interactions between different layers belonging to various scientific, technological, social and human domains at local and global scale. It is impossible to think of tools embedded in a rigid architecture implementing an approach top-down to face all the barriers and constraints of disaster resilience. The solutions must be based on methodologies and standards applied with a certain degree of flexibility that takes into account the local unique characteristics and history, but with a mid/long-term perspectives strategy. Any evolution of data and knowledge systems towards an enhanced digitized public good management, including cultural and natural heritage, should start from existing conditions and available resources in order to implement sustainable and feasible actions for resilience. The main general constraints in operationalizing data and knowledge are on data access and completeness, analytical challenges, human and technological capacity gaps, bottlenecks in coordination, communication, and self-organization. With the DMF it is possible to identify the gaps in terms of data access and completeness whereas with the principles, the methodologies and the standardized approaches described in this document and tested in the WP1 it is possible to overcome the other constraints concerning capacity, organization and coordination. Anyway, the effectiveness of the DMF is strictly linked to the local stakeholder requirements and the risk and resilience indicators. Indeed, the datasets themselves acquire sense only if filtered and correlated to the real needs of the OLs that are continuously changing and evolving.

7.1 Gaps

The analysis of the DMF compared with the GLOCAL requirements and the table of indicators depicts a complex and heterogeneous scenario about data and knowledge source identification. Most of the identified data fit with the requirements and, in some way, are available to generate the indicators mentioned in the list developed within T2.2, even though the list is not yet consolidated and validated at the time of writing this document.

There are still some gaps concerning the real accessibility of the identified data that can limit their usability for the SHELTER purposes.

Concerning new data that will be produced during the project, the DMF reports only a limited list of potential datasets since the most part of new data will be clearly described in detail once the tools will be defined, designed and implemented to match with the consolidated and verified requirements and indicators.

Global data have a high level of accessibility, reliability, interoperability, most of them are in machine readable and standard formats, they are uploaded regularly and well documented, but due to the large variety of these data, to be operated they need skills and tools. For example, the Copernicus initiative is a big European programme put in place 6 years ago that provides free and open data, but it is still broadly unknown outside the geoscience community. Despite the high level of accessibility and interoperability of the provided data, the transfer of this unprecedented amount of information encounters skills and technological barriers that can be overcome only with a strong effort in training and education programmes. The Copernicus initiative offers many benefits, but the most relevant is that this programme is continuously improving in terms of number and quality of the provided datasets and new examples of product and services are presented every year that are worth to be carefully monitored.

IoT is a relatively new concept that is becoming popular in recent years thanks to the diffusion of broadband infrastructures that can connect devices and sensors. Despite their diffusion IoT data sources are not exempt from weaknesses. Various factors may cause the IoT monitoring system unreliable: first, security and unified standards are not always guaranteed; second, as a consequence of the local scale characterizing IoT data, the languages used for the related metadata, instruments instructions and technical definitions are not homogeneous; finally the quality criteria are often unharmonised.

Within SHELTER a clear distinction must be done between existing IoT sensors and new installed IoT sensors. Existing IoT sensors, devices, sensor networks and local monitoring systems have been installed with a large variety of technologies, standards, architectures that need to be translated and homogenised by means of dedicated connectors to be integrated into the Data Driven Platform.

For what concerns the new sensors that will be installed locally during the project, it is necessary to design the architecture of the installation and of the monitoring system to be compliant to the common standards of data access and interoperability. As previously described (see Section 4.3), the technical referents of each OL, with the support of SHELTER technical partners, will be responsible for the installation of the new IoT sensors. After the end of the project, the OLs will take on the management of the newly installed sensors.

Data ownership is a common (existing and new sensors) issue that needs to be properly managed because it is not only a matter of legal aspects (license), but also a matter of practical aspects concerning authorized requests to retrieve data.

Within SHELTER, managing the IoT data layer means managing a combination of sensors, standards, systems, network of sensors that need to be integrated together trying to fill the technological gaps whenever is possible.

Crowdsourcing and social media

The crowdsourcing and social media data sources have been considered as useful source of information for disaster resilience monitoring only in recent years, thanks to the diffusion of telecommunication personal devices and the growth of telecommunication infrastructures that guarantee an incredible broadband capacity.

Crowdsourcing and social media present two main categories of issues:

- Technical issues concern the availability, the reliability, the accuracy, the accessibility, the exploitation for commercial and research purposes, the fact that the data owners are often private ICT companies that own all the rights and that can unilaterally change the terms and conditions of the data.
- Ethical issues concern again the ownership of the data, the personal data protection of the users that should be aware about the use that will be done with the information and the data which, more or less unconsciously, are being shared.

Dealing with crowdsourcing and social media data implies a strong and massive implementation of AI solutions to filter, organize and correct the bias of the data and also to extract the meta information like for example trends, population density, population movements to be used in DRM both as early warning system and for post disaster recovery monitoring.

Socio-economic data sources

The main socio-economic data sources are the national and international statistical portals that can provide most of the statistical indicators concerning population, economy, etc. mentioned in the indicators table.
The international organizations like World bank, United Nation, FAO, UNESCO and OECD already provide large amounts of well-structured data with metadata and related reports and documents in different languages that allow full access and exploitation also for external experts. The data are also exposed in machine-readable format facilitating the interoperability with data analysis tools. The national statistical departments expose data which are often in local language only and to be fully exploited and integrated with other datasets they need to be interpreted and translated.
Statistical data might result unclear to non-statistical experts and to be fully exploited and integrated with other data they need to be further elaborated, filtered and synthesized.

Local data sources condition is well described in Chapter 4. The presented graphs demonstrate that there is a large heterogeneity of readiness of the systems and the data due to the history of different countries in dealing with the resilience of natural and cultural heritage and with advanced geospatial data infrastructures implementation.

A large heterogeneity implies several technological issues briefly listed below:

- Local data standards
- Local data access

- Local data maturity
- Local data reliability
- Local data technologies
- Local data in local language
- Local data integration

The detailed description of these issues is provided in section 4.6, but, more in general, besides the listed issues, there are some further gaps to be mentioned which regard the full awareness of the importance of having a **long term data management strategy** and the **completeness and representativeness** of the local data identified so far.

The first point is quite clear, and it deals with the process of assimilation of the new concepts regarding big data and digital data era at local level. This process might take time and it needs a strong support and a visionary approach of the local stakeholder and the local governments that have to be convinced about the importance of implementing a data driven decision making methodology.

The second point deals with the data identification and collection process which must consider a project-life long process since the DMF needs to be continuously reviewed and updated whenever new inputs emerge, new data are generated, new indicators are reviewed and validated. This means that new data will be added, and some existing data will be dismissed since they could be considered obsolete or no more relevant for the scope of the project.

General considerations concerning the gaps of all kind of datasets and data sources investigated so far can be summarized in:

- Knowledge in data management strategies and standards
- Knowledge of tools to support the data management actions
- Awareness at all the levels of natural and cultural heritage management and government scale of the relevance of a data management strategy
- Refresh of identified data. Dataset and data strategy can expire rapidly for many reasons. New ICT solution and device diffusion and new environmental trend due to climate change can turn upside down the hazards prioritization list, making necessary the implementation of new parameters monitoring strategy.
- Completeness of dataset and data source identified so far. The DMF represents the state of the art of identified data at this state of the project, but we are aware that further datasets and data sources will emerge during the life cycle of the project thanks to the continuous interaction with the OLs stakeholder and the solutions implemented by the technical partners as response to the user requirements and the resilience and disaster risk management indicators that have been identified and that will be validated in the next steps.
- In the interaction with the Open Lab of Seferihisar it has emerged a relevant lack of local data that needs to be solved.

- For those datasets that are available and accessible, a sort of domain-oriented framework exposing the link between datasets to be used for resilience and DRM purposes looks often necessary to fully exploit the data sources. The data models that are going to be developed in SHELTER can provide this framework.

In the next section, some suggestions and mitigation actions are indicated to fill the above mentioned gaps.

7.2 Roadmap

The content of this section partially reflects the considerations described in the previous section and defines the strategy and the next actions to be implemented both within SHELTER and beyond the project.

The proposed roadmap can be described in three main pillars: the DMF update, the data management knowledge transfer and the data management SHELTER tools implementation.

7.2.1 Knowledge transfer within SHELTER consortium

The technical complexity of SHELTER data and SHELTER analytical tools could make the stakeholders feel like they represent yet another technocratic approach to development. However, the open and participatory nature implemented in SHELTER could increase the trustworthiness of data sources.

One of the most effective actions that can be taken in SHELTER deals with education, training and knowledge transfer. This action can return immediate effects and benefits allowing local stakeholder to design and implement improvement and extension of existing local data infrastructure as well as to exploit existing and mature solutions already implemented in other similar cases.

The principles of data management are well explained both in this document and in the Data Management Plan and they can be considered and used as tutorials to be disseminated within the consortium.

The DMP provides principles and references to delve into the data management concepts and it also reports links to data management tools like open repositories, guidelines, license definitions etc. The DMP has been considered as the baseline to design the DMF as described in section 3. Anyway the DMF building process can be considered as a practical example of knowledge transfer. Indeed, thanks to the interactions with the SHELTER partners, concepts about data management have been exchanged, elaborated and extended with the partners in order to create a tool in which both FAIR principles and SHELTER needs could converge into a unique, flexible and scalable tool.

After the building process, the DMF has been compiled with many interactions with technical partners and Open Labs. Detailed instructions reporting not only how to compile the fields, but also the ratio behind each field have been provided and many bilateral interactive sections have been made to support the datasets identification.

The next steps within the project will be to keep interacting with the other partners by promoting the usage of DMF and the other tools developed in WP1 and by presenting the content and the guidelines defined in this document especially to the OLs during the periodic OLs workshops.

Apart from the dissemination and training, other specific knowledge transfer actions will be put in place, based on the particular need of OLs and partners. This is the example of the Sava River Basin International Commission (ISRBC) case in which a mock-up of cultural heritage sites inventory data model template has been designed in response to a specific request of support. In this case the technical and scientific knowledge of WP1 partners has been '*operationalised*' to provide support to design a data management tool in compliance with the FAIR standards.

Moreover, through the future interactions with the OLs, the data gathering process for the Intangible Heritage will be started. As highlighted in section 3, in fact, this kind of knowledge is particularly hard to retrieve and to describe. Consequently, it will be necessary to define, in collaboration with POLITO, a specific protocol, taking into account the necessary evolution of the DMF structure.

For what concerns new data identification and gathering, future technologies involving IoT data, such as Edge Computing and Edge computing enabled devices, should be explored as they are growing in popularity. Terrestrial based 5G infrastructure rollout is fast approaching, as is a push towards Edge computing, which is the concept of processing and analyzing data in servers closer to the applications they serve. This can drastically change the technology landscape, starting with Edge enabled devices (sensors, probes, etc.) and ending with Cloud infrastructure. It also flows towards a meso to macro risk management and response governance model based on data processing locality and monitoring.

The concept of Edge computing may not be revolutionary, but the implementations will be. These implementations will solve many growing issues including reducing energy use by large data centers, improving security of private data, enabling failsafe solutions, reducing information storage and communication costs, and creating new applications via lower latency capabilities.

7.2.2 The Data Mapping Form in SHELTER

As repeated many times in this document, the DMF represents the most added value output of T1.1. It must be considered both an analytical tool and a framework to organize the datasets tailored to SHELTER purposes. Indeed, the elaboration and the analysis presented in section 4 have been made possible only by a data framework, the DMF, duly compiled with detailed information.

In this sense it is clear that for the rest of the project the DMF needs to be continuously updated with missing datasets like those concerning Seferihisar OL and with the new datasets generated by the technical partners of SHELTER that are in charge to provide the technological solutions in response to the user requirements. Nevertheless the DMF must be updated in compliance with the methodology described in section 3 that implies a continuous revision of the identified datasets with respect to the consolidated and validated indicators and criteria described in the DMP. This means that during the project life cycle it might happen that a newly identified dataset of better quality replaces a similar one or that a dataset might become obsolete or no more relevant due to new requirements emerging from the interaction with the OLs stakeholders.

By keeping updated the DMF content with datasets as much as possible compliant with the standards and the good data management practices including the replicability and the interoperability, the Data Lake and the Data Driven Platform will become fully operative.

The DMF is subject to continuous updates not only in terms of contents, namely the newly identified datasets, but also in terms of data structure. As reported in section 3.1, the Phase 4 of the DMF methodology deals with the approach to evolve the data structure by adding new criteria, new parameters in response to the feedback and the inputs that come from the OLs workshops and from the interaction and the output of the other tasks. The example is the tool matrix sheet mentioned in section 3.1. It is clear that a duly compiled data structure is particularly useful to associate tools, concepts, approaches, techniques and solutions with identified datasets in order to make them quickly applicable. The DMF content can be also directly transferred to the Data Driven Platform as metadata description associated with each identified dataset. In this way the Data Mapping Form conceived to provide a detailed description of the identified datasets can evolve into a Data Management Framework to create the connections between data and the rest of the output generated in SHELTER.

The DMF can be used as an analytical tool, as already demonstrated in section 4.6. A further example of data analysis is depicted below:

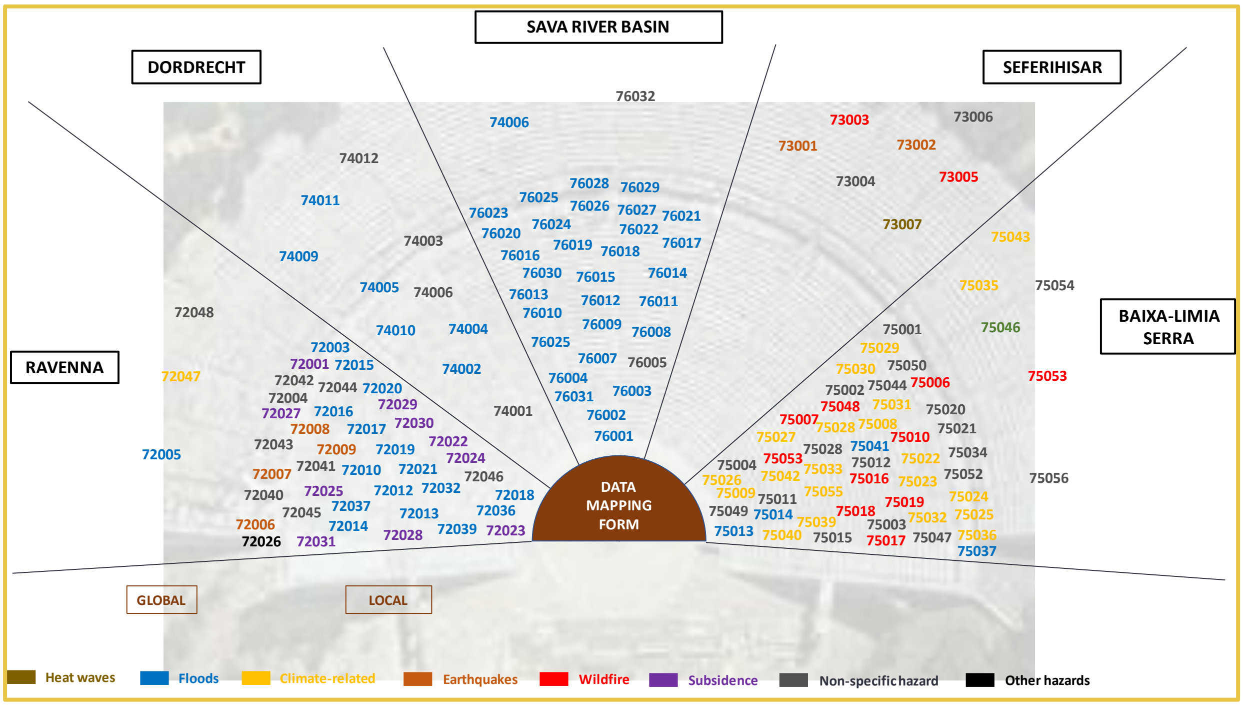


Figure 18 – Datasets distribution visualization in terms of addressed hazard, scale (global/local) and Open Labs

This figure represents the distribution of the datasets indicated as dataset ID per Scale (global/local), per OLs and per type of Hazard. By means of this kind of graphic visualization it is quite easy to characterize the datasets. Indeed, as already mentioned in this report, Seferihisar is the only OL with no local data identified so far.

7.2.3 The data management tools

As aforementioned, the data must not be considered as a panacea. Without a data framework, data science knowledge, advanced exploitation and analytical tools, the data might result useless or even damaging due to the risk of confusion and misleading effects. Data management tools are as important as the data themselves and they must be designed and developed to extract the maximum value from the data.

Furthermore, data and knowledge driven systems should be flexible and easy to adapt themselves along the way whenever any change or integration occurs.

In the digital era, besides the big data challenges there is the great challenge of developing semantic layers that can make the big data fully exploitable by Artificial Intelligence (AI) tools based on machine learning (ML) techniques. Making a dataset compliant to machine readable data formats means making an intelligent use of the tags and labels associated with metadata.

To this scope SHELTER has given a great relevance to the ontology which has been developed in Task 1.2 and that it is briefly described in section 5.1. To be effective, the ontology needs to be shared and used at all levels of the SHELTER project. For this reason, in the framework of T1.2 activities, a **SHELTER WIKI** [AD.2] exposing the ontology content has been proposed and implemented. The WIKI, as well as the DMF, needs to be continuously updated with the inputs collected from the partners and the interaction with the OLs. A multi language version, at least of the main concepts and definitions, would be extremely useful to allow a more fruitful knowledge exchange and transfer even beyond the SHELTER project.

Making the data machine-readable is a fundamental step to advance the digitization process towards the full implementation of the FAIR principles. The analysis of the datasets identified in T1.1 and T1.2 as beforementioned, depicted a very heterogeneous and complex scenario concerning the variety of data and knowledge to be operationalized within SHELTER. To manage such an expected amount of various data, the concept of **Data Lake** has been introduced.

The Data Lake is subject of T1.3, but it is strictly linked to T1.1, T1.2 and to the DMF. A data lake is a data repository that stores a large and varied amount of structured and unstructured data. The data lake is data agnostic and general purpose enough to be able to store all kinds of outputs identified in T1.1 and T1.2 simply by making use of tags.

Since the Data Lake is a fundamental component of the Data Driven Platform, it is crucial that the design and the implementation of this component reflects the variety and the features of the dataset described in the DMF and of the knowledge described in D1.2.

To this end, the data samples associated with the datasets described into the DMF will result extremely useful to implement not only an effective Data Lake, but also for all the other data driven components that will deal with data.

The same crucial benefit of the DMF is envisaged for the Data Models. As described in section 3.2, the **SHELTER's multiscale data model** will represent the relevant information to characterize the Resilience ID. It is therefore a geospatial data model with information at different scales based on international standards such as CityGML and INSPIRE. This means that the data model, to be effective, must include all the kind of geospatial data described in the DMF. The Data model can be seen as a tool to interpret the data for a specific scope; in the case of SHELTER the data model is expected to tighten up the data and the knowledge from different scientific, technical and social domains for cultural and natural heritage safeguarding. Without a properly designed data model neither the identified data nor the data access or data exploitation tools like the data lake, make sense.

7.2.4 The example of the Rapid Damage Assessment technology

The Rapid Damage Assessment module (RDA) is responsible for providing satellite mapping for both burned and flooded areas. A diagram of the architecture is shown in Figure 19.

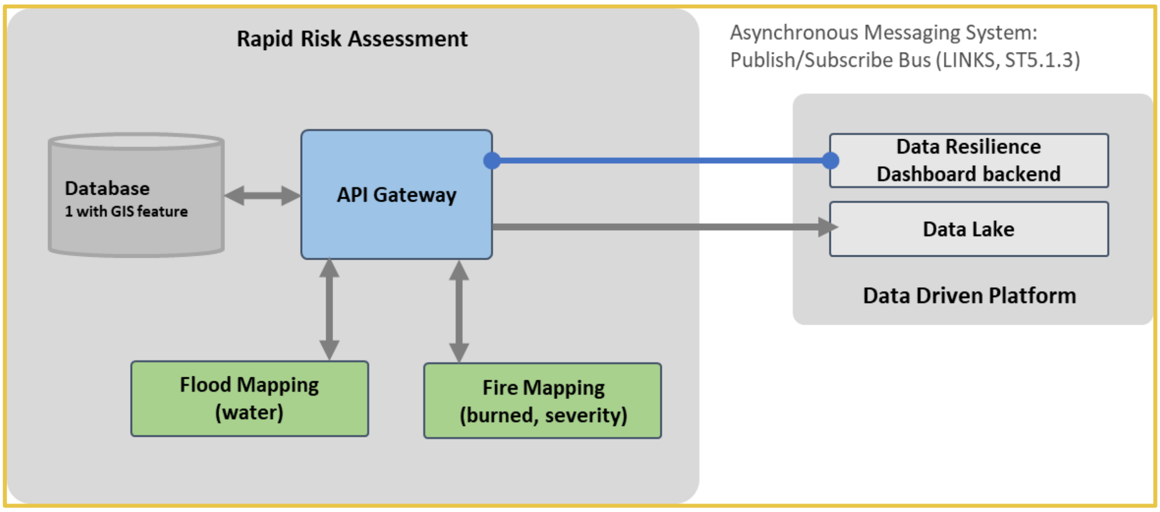


Figure 19 - Rapid Damage Assessment module architecture

The module will be accessible through the Data Resilience Dashboard (DRD), a web application which provides: *i)* information about data previously processed by the platform, and *ii)* the functionality to submit a new mapping request. The website is thought of as an interface to guarantee easy and prompt access to the end-users, but it will make use of an *API Gateway*, which allows it to be accessed by external systems. Every access or request sent to the Rapid Damage Assessment module must be authorized by a centralized authorization module, the *AA Server* (delivered by ST5.1.2).

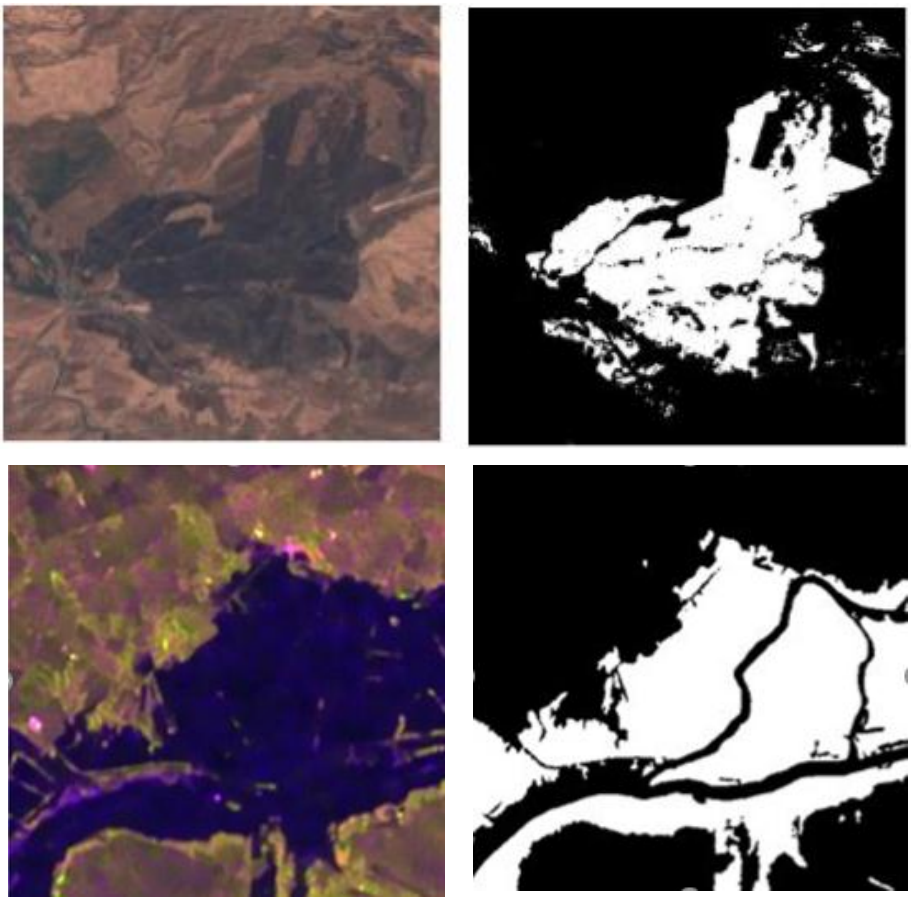
Map Requests for the RDA module will be sent through the *Asynchronous Messaging System* (AMS) (delivered by ST5.1.3) as a JSON file containing the following information:

- a *time_range* of interest
- a geographical *Area of Interest* (AoI)
- the type of mapping to be performed (*Task)*
- the *Hazard* type

The ASM will implement the Publish / Subscribe design pattern to allow the RDA module as well as any authorised users to be informed when there is a new Map Request and when the mapping result is available. From the *time_range* and the *AoI* the RDA module will be able to: fetch relevant satellite data (from Copernicus Sentinel), process the acquisitions, run the machine learning models, and return a sequence of maps, which will be delivered to the SHELTER Data Lake (delivered by T1.3 and ST5.1.1) in a format that will enable the creation of map layers that could be visualized by the DRD user. The *Hazard* type indicates the event to consider during the mapping phase, which can be *Fire* or *Flood.* Finally, the *Task* indicates whether to compute a *delineation map* or a *grading map.*

The *deployed Machine Learning models* will handle the request and produce the specified mapping. Acquisitions inside the specified time-range will be pre-filtered to exclude areas fully covered by clouds. In both *Wildfire* and *Flood* mapping, the computation of a *delineation map* can be requested. A delineation map is an image which highlights the area/s affected by the event, namely flooded or burned areas for the flood and wildfire case, respectively. Each pixel of the satellite acquisition will be marked affected or not. Furthermore, for wildfire events, the *Wildfire Mapping* module will compute an estimation of the damage severity in the affected areas, which is referred to as a *grading map*. The grading map is similar to the delineation map, but it will contain the following classes: *Negligible to slight damage*, *Moderately Damaged*, *Highly damaged*, or *Completely destroyed*. An example of delineation and grading maps is shown in Figure 20.

(a) *(d)*



(b) *I*

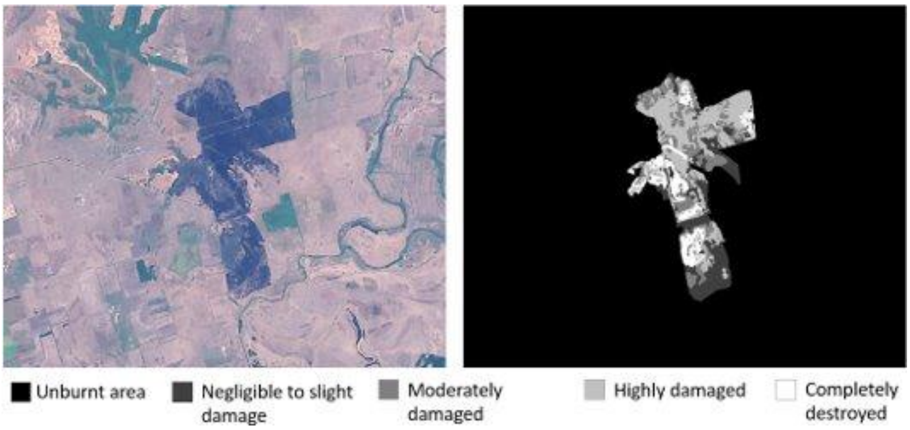
(c)  (f)

Figure 20 - Example of regions affected by hazardous events (a, b, c) and the related delineation (d, e) and grading (f) maps

The outcomes of the ML models will be GeoTIFF files, containing information about the acquisition date, the AoI, and the map itself. After the computation is over, a notification will be sent to the AMS to inform the interested users that the mapping is available.

The Data Lake component will serve as a global file repository, to store the satellite images and the mapping results. Finally, the Database will record all structured information about mapping requests and results, file source locations (URLs to the Data Lake), and logs to monitor the correct functioning of the RMA submodules.

8 Conclusion

Climate change will in all likelihood increase the frequency and intensity of hydrometeorological hazards in varying complex ways that are expected to only worsen the situation. This confluence of factors has led to increase the calls dedicated to make disaster risk reduction a core development concern, as well as to promote an understanding that disaster risk reduction is a development investment. The cultural and natural heritage is exposed to these risks and it is necessary to implement a data driven approach to increase the climate resilience and to protect them by monitoring and detecting hazards, mitigating their effects, and assisting in relief efforts. Building and reinforcing the resilience means that vulnerable communities and countries as complex human ecosystems become able to adapt and to maintain equilibrium in the face of natural hazards.

As reported in '*The European Green Deal*' the Accessible and interoperable data are at the heart of data-driven innovation dealing with climate change adaptation and mitigation. Despite this great value recognized to the data that the EU data policy *('A European Strategy for data')* has been defined as the center of the digital transformation of the economy and society, the same report mentions some crucial problems that in SHELTER are well known:

- Availability of the data
- Data Interoperability and Quality
- Data Governance
- Data Infrastructure
- Skills and data literacy

These problems have been seriously considered since the beginning of the project, particularly in WP1. Indeed, the "…*application of standard and shared compatible formats and protocols for gathering and processing data from different sources in a coherent and interoperable manner across sectors and vertical markets should be encouraged through the rolling plan for ICT standardisation …*" is the one of the main principles that have been applied in the design and implementation of the methodologies and tools in WP1. Scope of WP1 is also to promote and share all the knowledge, the guidelines and the approaches to mitigate the barriers listed above.

Despite the great benefit expected from the diffusion of big data culture (data sciences, tools, devices and infrastructures), it has to be clarified that data is not a panacea. However, as emerged by the analysis described in this report, in the digital era the data can strongly support the process of disaster risk management in many different ways:

- Monitoring the hazards
- Assessing exposure and vulnerability
- Supporting the disaster risk management

- Assessing the resilience of natural and cultural heritage
- Engaging the local stakeholders and communities

As beforementioned, data only cannot be considered exhaustive and self- standing but rather as part of an ecosystem implementing an approach made up of three main pillars: **the data**, **the tools** to collect, produce, access and analyze the data, and skilled, trained **expert people** with the capacity of properly collecting and analyzing the data and generating those added value information useful to support the DRM phases.

The variety of identified data demonstrates how it can be complex to deal with cultural and natural heritage resilience and how it can be hard to implement an operational DRM approach. Leveraging a growing amount of data requires navigating and linking highly complex technological, political, and socio-economic systems. In SHELTER all the resources in terms of capacity, knowledge and tools are available to make the full exploitation of the data.

By means of the datasets described into the DMF, the indicators identified in T2.2 can be directly or indirectly produced by the extraction from the data sources or by computation performed by SHELTER tools making use of those dataset. The expertise and the experience within SHELTER allow the transfer of knowledge and the application of these indicators for building strategic decisions support systems and HA resilience dashboards.

The next steps in SHELTER should be focused on filling the gap, integrating the new dataset generated in other WPs, keeping updated the DMF to feed the Data Lake, the Multiscale and Multisource Data Model and the other data components of the Data Driven Platform to operationalize the data and the knowledge.

The strategy and the approach proposed in this document aims also at bringing the data beyond SHELTER. The implementation of the data management plan together with the actions of dissemination and knowledge transfer of data management concepts will reinforce the accessibility, replicability and reusability of the datasets and the knowledge ensuring their operationalization for long after the end of the project.

9 References

Scientific publications:

Ref	Name of document/link
[1]	Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016) https://doi.org/10.1038/sdata.2016.18
[2]	Sablier, M., Garrigues, P. Cultural heritage and its environment: an issue of interest for Environmental Science and Pollution Research. Environ Sci Pollut Res 21, 5769–5773 (2014) https://doi.org/10.1007/s11356-013-2458-3
[3]	Stefanidis, A., Crooks, A. and Radzikowski, J. Harvesting ambient geospatial information from social media feeds. GeoJournal 78, 319–338 (2013) https://doi.org/10.1007/s10708-011-9438-2
[4]	Middleton, S.E., Middleton, L. and Modafferi, S. Real-Time Crisis Mapping of Natural Disasters Using Social Media IEEE Intelligent Systems, vol. 29, no. 2, pp. 9-17, Mar.-Apr. (2014) https://doi:10.1109/MIS.2013.126
[5]	Nguyen, Duc T. and Jung, Jai E. Real-time event detection for online behavioral analysis of big social data Future Generation Computer Systems, vol. 66, pp. 137-145 (2017) https://doi.org/10.1016/j.future.2016.04.012
[6]	Chae, J.,Thom, D.,Jang, Y., Kim, S., Ertl, T.,Ebert, D.S. Public behavior response analysis in disaster events utilizing visual analytics of microblog data Computers & Graphics, vol. 38, pp. 51-60 (2014) https://doi.org/10.1016/j.cag.2013.10.008
[7]	Dwivedi, S.K. and Ganesh, C. A Survey on Cross Language Information Retrieval. International Journal on Cybernetics & Informatics. 5. 127-142 (2016) https://doi.org/10.5121/ijci.2016.5113
[8]	Zarnić, R., Rajcic V., Vodopivec, B. Data Collection for Estimation of Resilience of Cultural Heritage Assets. In: Ioannides, M., Magnenat-Thalmann, N., Papagiannakis, G. (eds) Mixed Reality and Gamification for Cultural Heritage. Springer, Cham. (2017) https://doi.org/10.1007/978-3-319-49607-8_11
[9]	Davis E. and Heravi B., Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation. Journal on Computing and Cultural Heritage, vol. 14, Issue 2, pp. 1-14 (2021) https://doi.org/10.1145/3429458

Reports and other documents:

Ref	Name of document/link
[REP1]	Corporate author(s): Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (European Commission), PwC. Copernicus services in support to Cultural Heritage report. EU Publications (2019) https://doi.org/10.2873/795267
[REP2]	United Nations – Office for Outer Space Affairs. UN recommended practice for flood mapping. Resolution 61/110 of 14 December 2006. Available at: http://www.un-spider.org/advisory-support/recommended-practices/recommended-practice-flood-mapping/in-detail

[REP3]	Eurostat environment overview. Available at: https://ec.europa.eu/eurostat/web/environment
[REP4]	Rogushina, J., Gladun, A., Прийма, С.М., Прийма, С.Н., & Pryima, S., Use of Ontologies for Metadata Records Analysis in Big Data, (2018) CEUR Workshop Proceedings; Vol. 2318 (p.46-63). Available at: http://elar.tsatu.edu.ua/handle/123456789/7431
[REP5]	Ostermann and Granell, National Academies of Sciences, Engineering, and Medicine, Reproducibility and Replicability in Science, Washington, DC: The National Academies Press (2019). Available at: https://doi.org/10.17226/25303
[REP6]	European Commission, The European Green Deal, Brussels, COM(2019) 640 final, (2019). Available at: https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf
[REP7]	European Commission, A European Strategy for data, Brussels, COM(2020) 66 final, (2020). Available at: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

Websites:

Ref	Name of website/link
[URL1]	Data Mapping Form
[URL2]	International Sava River Basin Commission (ISRBC), Sava GIS Geoportal. Accessible at: http://savagis.org/map;jsessionid=366BE4DD718DB7BF897798BF8C18251E
[URL3]	Centre for Research on the Epidemiology of Disasters (CRED) EM-DAT – The International Disaster Database. Accessible at: https://www.emdat.be/
[URL4]	Climdex portal. Accessible at: https://www.climdex.org/
[URL5]	European Commission, The Copernicus Programme. Accessible at: https://www.copernicus.eu/it
[URL6]	Copernicus Emergency Management System (EMS). Accessible at: https://emergency.copernicus.eu/
[URL7]	Statistical Office of the European Communities. EUROSTAT: Regional statistics: Reference guide. Luxembourg, Eurostat, (1990). Accessible at: https://ec.europa.eu/eurostat/web/main/home.
[URL8]	Munich RE – Munich Reinsurance Company. Accessible at: https://www.munichre.com/en.html
[URL9]	Istat - Istituto Nazionale di Statistica. Accessible at: https://www.istat.it/en/
[URL10]	TurkStat – Turkish Statistical Institute. Accessible at: https://www.tuik.gov.tr/
[URL11]	Ministry of Culture and Tourism – Republic of Turkey. Accessible at: https://www.ktb.gov.tr/?_dil=2
[URL12]	StatLine – Statistics Netherlands' database. Accessible at: https://opendata.cbs.nl/statline/#/CBS/en/
[URL13]	INĒ – Instituto Nacional de Estadística. Accessible at: https://www.ine.es/en/
[URL14]	Instituto de Estudos do Territorio (IET), Conselerìa de Medio Ambiente, Territorio e Infraestruturas, Información Xeográfica de Galicia - Xunta de Galicia, WMS services available at: http://mapas.xunta.gal/servizos-wms
[URL15]	Rijksdienst voor cultureel erfgoef (RCE), Nationaal Georegister – NGR. Accessible at: https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home
[URL16]	Agenzia Prevenzione ambiente energia Emilia-Romagna (ARPAE). Accessible at: https://www.arpae.it/it

[URL17] Pasqua Recchia, A., Capponi G. (Ministero dei Beni e delle Attività Culturali e del Turismo), Moro, L., Cacace, C. (Istituto Superiore per la Conservazione ed il Restauro), Vincoli in Rete. Accessible at: http://vincoliinrete.beniculturali.it/VincoliInRete/vir/utente/login

10 Annexes

10.1 Annex I: Data description template applied in SHELTER, related instructions and supplementary documents

The complete version of the Data Mapping Form is available here

Seferihisar	Description	Existing/Foreseen data	Type	IoT	Format	Size	Time coverage	Area coverage	Spatial resolution/scale	Temporal resolution	Update frequency	Collection	Licence	Ownership/author	Access mode	Access restrictions	Access links	Metadata	Storage	Processing	Analysis	Tools	Application field	Hazard type	DRM phase	End-users	Future data	Comments	Data example
73001																													
73002																													
73003																													
73004																													
73005																													

Figure A- 1 – Data Mapping Form template (Seferihisar OL example)

Examples Filename convention

Field	Description	Example value	Comment
Field separator	separator	_	
dataID	the unique data ID used in the data Mapping Form	72001	First digit = WP Second digit = Task Third digit = eventual subTask Last 3 or 2 digits = progressive number
owner	owner short name	UNIBO	the partner responsible for the data (refer to the Partners_short_names.docx on SharePoint: https://tecnalia365.sharepoint.com/:w:/r/sites/t.extranet/sp070767/Shared%20Documents/SHELTER/WP1/T1.1/DataExamples/Partners_short_names.docx?d=w6eac2739437346dd9f70388467abfbda&csf=1&e=s1fnfR)
description	20 character description of the data	subsidence_spd_iso	
dateStart	minimum validity/acquisition datetime contained in the data ISO 8160, UTC timezone	20200107T112707Z	https://en.wikipedia.org/wiki/ISO_8601 if data has daily resolution, put time at 000000Z. If data has monthly resolution put 01 as day, if yearly resolution put 01 as month
dateEnd	maximum validity/acquisition datetime contained in the data ISO 8160	20200110T120000Z	https://en.wikipedia.org/wiki/ISO_8601 if data has daily resolution, put time at 000000Z. If data has monthly resolution put 01 as day, if yearly resolution put 01 as month
country	2 digit country code as per ISO_3166-1_alpha-2	IT	https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2
extension	standard file extension (file format)	.csv	https://en.wikipedia.org/wiki/List_of_file_formats
filename	complete filename of this example	72001_UNIBO_subsidence_spd_iso_20200107T112707Z_20200110T120000Z_IT.csv	

Figure A- 2 – Data Mapping Form, example files naming convention

Instructions:
1. Go to the sheet relative to the Open Lab you are involved in or, if you are a technology partner, use the sheet 'Template' renaming it with 'P_*your Shelter organisation number*' (e.g. P_15 for SISTEMA)
2. Fill in the empty cells (one row per data) according to the instructions below (each colour is related to a group of steps in the data value chain). **Only for technology partners**: compile the form exclusively for already existing dataset or foreseen dataset with already defined features, otherwise wait until the user requirements are defined.
3. Once you have finished, please notify via email to: quarta@meeo.it (with folegani@meeo.it in Cc) and to claudio.rossi@linksfoundation.com

Task/Subtask (only for technology partners):
Indicate the ID of the related Task
Input task (only for technology partners):
Indicate the ID of the Task that will provide input for producing the described dataset
Input direction (only for technology partners):
Choose one of the proposed symbols (--> indicates that input are going from one task to another in one-way direction; <--> indicates that input task are providing information for the reference task and it is also receiving from it)
Data description:
Shortly describe your data
Existing/Foreseen data:
Use 'Existing' for data already existing before the start of project activities and "Foreseen" for data that is planned to produce during the development of the project
Data type:
Which kind of data do you have?
Define the data you have according to the following classification:
a. **Structured data**: any data that resides in a fixed field within a record or file data contained in relational databases, spreadsheets, and data in forms of events such as sensor data;
b. **Semi-structured data**: tags or other types of markers are used to identify certain elements within the data, but the data doesn't have a rigid structure (e.g. in object oriented databases one often finds semi-structured data);
c. **Unstructured data**: data lacking of a pre-defined data model (e.g. books, journals, images, documents, metadata…)
d. **New generation Big Data**: semi-structured and unstructured data, often in combination with structured data:
- sensor data (IoT data from in-situ sensors and telemetric stations);
- imagery data from unmanned aerial sensing platforms (drones);
- imagery from hand-held or mounted optical sensors.
Is it your data a geospatial or a non-geospatial one?
Geospatial data is information that has a geographic aspect to it (the records have coordinates, address, city, postal code or zip code included with them (e.g. rasters, vectors..)
IoT:
Choose "**yes**" or "**no**" to indicate whether the described dataset derives from IoT sources (i.e. data from a network of devices and physical objects that can connect to the Internet, recognize other devices and objects and communicate with them)

Data format:
Indicate the format of your data (e.g. file with its extension, hard support - i.e. paper - report, bulletin, web service, spreadsheet...)
Data size:
If you have such information, insert the size occupied by your data when stored. This can refer either to the total data size or to the size relative to 1 year of data. Indicate which of these 2 cases are you referring to.
Data time coverage:
Indicate the available time range of your data, both for archive and for forecast data. Provide the lead time in case of forecast data. In case of unstructured data (e.g. a report), indicate the validity time of the data.
Data area coverage:
Indicate the geographic extension of your data (global, regional, local or Europe, Asia...). In case of unstructured data (e.g. a report), indicate the area which the data is applied to.
Data spatial resolution/scale:
If you have any geospatial data, indicate what is the relative spatial resolution (for rasters) or scale (for vectors)
Data temporal resolution:
Indicate the amount of time needed to revisit and acquire data for the exact same location
Data update frequency:
Indicate how often your data is updated
Data collection:
Who has been collected your data?
Data ownership/author:
Indicate ownership and/or author of your data and if it results from other projects outcomes
Data licence:
Indicate what kind of licence characterize your data (e.g. Creative Commons,...)

Data access mode:
How is your data accessible?
Describe how is your data accessible (through API, GUI, WMS, WFS...). In case of access through API indicate the relative documentation link

Data access restrictions:
Does your data have open access?
Indicate if your data can be accessible without restrictions from copyright, patents or other mechanisms of control.

Data access links:
Provide any link to websites/portals that can give access to your data with the relative access credentials (username, password)

Metadata:
Does your data come with metadata?
If yes, indicate if the metadata is compliant with any particular standard (e.g. INSPIRE). If not, leave the cell empty.

Data storage:
How and where is your data stored?
Indicate if your data is stored on a cloud or on-premise servers

Data processing:
Indicate if and how your data has to be processed or organised for analysis. For instance, these may involve placing data into rows and columns in a table format (i.e., structured data) for further analysis, such as within a spreadsheet or statistical software

Data analysis:
How would you like to exploit your data?
Describe which kind of analysis would you like to perform starting from your data (modeling, aggregation, prescriptive, predictive, descriptive, textual...)

Data tools:
What tool do you use?
Describe the tool you use to perform your data processing and/or analysis

Data application field:
What is the application field of your data?
List 3-4 keywords to describe the application field of your data (e.g. risk management, hazard monitoring, GIS, damages, marine, ..)

Hazard type:
Which one of the following hazard is your data related to?
Earthquakes - Storms - Floods - Heat waves - Wildfire - Subsidence - Climate-related - Non-specific hazard - Other hazards - Multihazards
If the 'Multihazards' option has been selected, please specificy which of the abovementioned hazards by using the 'Comments' field.
If the 'Other hazards' option has been selected, please add the related description in the 'Comments' field

Disaster Risk Management (DRM) phase:
Which of the following DRM phase your data is applicable to?
- **Prevention** (regulatory and physical measures to reduce the likelihood of a disaster event occurring or the severity of an event should it eventuate)
- **Preparedness** (measures taken to prepare for and reduce the effects of disasters)
- **Response** (the taking of appropriate measures to respond to an event)
- **Recovery** (the restoring or improving of livelihoods and health, as well as economic, physical, social, cultural and environmental assets, systems and activities, of a disaster-affected community or society).
In case your dataset is applicable to more than one step, select the 'Multiphases' option and use the 'Comments' field for specifying which of the abovementioned steps you are referring to.

Data end-users:
Who is going to use your data?

Future data:
Which kind of data are you missing and would you like to add to your existing data?

Comments:
Use this section to write any additional information you think is needed or to ask any questions

Data example:

Figure A- 3 – Data Mapping Form, compilation instructions

WP	Task	Tool/Solution Name	Description	Tool output format	Input from	Output module DDP	Partner	Visualization Output	prevention	preparedness	response	recovery	Seferihisar	Dordrecht	Baixa Limia-Serra	Ravenna	SAVA	data IDs	data example	Comments
3	3.5	Chatbot	Chatbot for crowdsourced data collection	web-based API	User	Any module using the API e.g. Data Resilience back-end	LINKS	List of data colelcted, aggregation with statistics, temporal graph (e.g. volume grouped by	v	v	v	v	v	v	v	v	v	35001	NA	data example will be available after the data interfaces will be defined
3	3.5	Social Media Engine	social media information extraction and event detection	web-based API	Twitter	Any module using the API e.g. Data Resilience back-end	LINKS	List of tweet, aggregation with statistics, temporal graph (e.g. volume grouped by topic) map		v	v	v	v	v	v	v	v	35002, 35003, 35004	NA	data example will be available after the data interfaces will be defined
3	3.2	Flood delination	provides on-demand delination of water from sentinel-1 . Can be used to map floods	geoTIFF	Data Resilience dashboard	Data Lake	LINKS	WMS Map Layer (2D)			v	v	v	v	v	v	v	32001	NA	mapping can be requested from web-based dashboard, and it will be pushed to the Data Lake. It will have to be imported
3	3.2	Fire delination	provides on-demand delination of burned area from sentinel-1&2 . Can be used to map burned areas	geoTIFF	Data Resilience dashboard	Data Lake	LINKS	WMS Map Layer (2D)			v	v	v	v	v	v	v	32002	NA	mapping can be requested from web-based dashboard, and it will be pushed to the Data Lake. It will have to be imported
5	5.4	Multi-Hazard Risk Assessment					EKO		v											
5	5.3	Resilience Index Assessment and Monitoring					GIS		v			v								
5	5.4	DSS for Adaptation and Recovery					TEC		v			v								
3	3.4	Prioritisation Matrix					NBK		v			v								
4	4.4	Maintenance Scheduler					EKO		v											
3	3.6	IMMERSITE					NBK		v		v	v								
3	3.1	Early Warning System					UMAS			v										
3	3.2	Rapid Damage Assessment									v									
3	3.4	Solutions Portfolio							v	v	v	v								
2	2.7	Adaptation Roadmap							v											
4	4.2	Risk Management Plan								v										
2	2.6	Agent Based Modelling								v	v									
4	4.2	Protocols for CH protection									v									
4	4.3	Backup 3D Models										v								
4	4.3	Recovery roadmap										v								

Figure A- 4 – Tool-Matrix

10.2 Annex II: the Data Mapping Form Evolution document

Data
Mapping Form
evolution

1 Scope of the document

The present document has to be considered as annex to the Data Mapping Form (WP1 Task 1.1) shared on SHELTER Sharepoint at the following link:

Data Mapping Form

Scope of the document is to collect future Data Mapping Form updates on the basis of the considerations derived from reviewing the already described dataset. The Data Mapping Form is in fact a living document and for this reason can be subject to further changes/improvements during the course of the project.

Any involved partners is allowed to give contributions, by indicating name of contributors and date. SISTEMA will evaluate the proposed change in coordination with WP5 partners and will integrate the new suggested update in case of positive feedback.

One of the following status definition has to be used:

a) *To be evaluated*
b) *To be integrated*
c) *Added*

1.1 Future suggested updates

- **SIS** (Marco Folegani, Maria Luisa Quarta) – 01/10/2019

  a) In the near future it would be useful to add to the data mapping form some keywords indicating which of the following data risk management phases could a certain dataset be related to (see T2.1): prevention – preparedness – response – recovery (*Added*)

  b) When a group of dataset is described under a unique data provider (e.g. Copernicus) the relative link should be moved to the field "Comment" and the single datasets need to be listed and described in details (with the relative link). These will allow to give more precise information about the dataset itself (e.g. data format, spatial and temporal resolution…). In any case the data provider link remains useful for the data platform. This implies that the data mapping form has to be updated constantly, also by performing regular check of the data links (*To be integrated*)

  c) During the course of the project it is suggested to add to the data mapping form further fields, which can be dedicated to the costs related to a given dataset (see Section 4.3 of D9.3_DataManagementPlan_V0.3) (*To be evaluated*)

  d) Each partner is suggested to add to the data mapping form the field "Post-it" to be used with internal purposes only as an extra section for taking note of eventual information considered useful, but that do not fit into the form (*To be evaluated*)

  e) During the course of the project certain dataset can become unavailable (for any reason) or it can be replaced by an updated version. In this case the relative metadata will always be accessible even if the underlying data is not or no longer available (*To be evaluated*)

  f) It could be useful to add a field to indicate whether it is possible to upload SHELTER output to a specific web portal, in order to contribute to the collection of useful data (e.g. Sava River basin has already a structured web portal). The field should contain the direct link for uploading or the wms link. This way the already existing users community can take advantage of new and updated datasets, sharing them with other users (*To be evaluated*)

  g) In order to link the Data Mapping Form to the DMP it is suggested to integrate the first one with a field containing specific keywords referring to the applied FAIR principles to the dataset in hand (e.g. 'Geospatial metadata' for 'Interoperability') (*To be evaluated*)

  h) To make a connection between already existing dataset and those that will be produced during the course of the project, it is suggested to add a field named "Related dataset" to be filled with the related dataset identifier (e.g. D01.D) (*To be evaluated*)

- **SIS** (Marco Folegani, Maria Luisa Quarta) – 07/11/2019

  i) To update the "*Hazard type*" Data Mapping Form field according to list of hazard reported in WP2 presentation (06/11/2019) (*Added*)

  j) To use the following data naming convention: <WP_number><Task_number><subTask_number><Progressive_number (2-3 digits depending on whether there is a subtask or not>, e.g. 35001 (*Added*)

  k) To specify which DDP module is producing a certain dataset (*Added – see Tool_Matrix sheet*)

10.3 Annex III: questionnaire for data info collection – WP6 requirements integration

Technology - Data types and format:

- ***Where does your data lies among the following types?***

  a. **Structured data**: any data that resides in a fixed field within a record or file data contained in relational databases, spreadsheets, and data in forms of events such as sensor data;
  b. **Semi-structured data**: tags or other types of markers are used to identify certain elements within the data, but the data doesn't have a rigid structure (e.g. in object oriented databases one often finds semi-structured data);
  c. **Unstructured data**: data lacking of a pre-defined data model (e.g. books, journals, images, documents, metadata…)
  d. **New generation Big Data**: semi-structured and unstructured data, often in combination with structured data:
     - sensor data (IoT data from in-situ sensors and telemetric stations);
     - imagery data from unmanned aerial sensing platforms (drones);
     - imagery from hand-held or mounted optical sensors.

- ***Is it your data a geospatial or a non-geospatial one?***

- ***If you have any geospatial data, what are the relative spatial and temporal resolution, format and what is the available time range?***

- ***How and where is your data stored?***

Data accessibility:

- ***In which format and in which way - open, restricted, hard support (i.e. paper) report, bulletin, web service, spreadsheet… - is your data accessible?***

- ***Does your data have open access?***

- ***Please, provide any link to websites/portals that can give access to your data***

Ethics and legal compliance:

- ***Who has been collected your data? What kind of licence and ownership characterize your data?***

- ***Does your data result from other projects outcomes?***

Data application field:

- ***What is the application field of your data? Please, list 3-4 keywords to describe your data***

- ***Who is going to use your data?***

- ***Which kind of data are you missing and would you like to add to your existing data?***

10.4 Annex IV: Survey for collecting information about IoT sensors and the related data

Collection of information about IoT sensors and the related data
Requested by: RED, POLITO, LINKS

Definitions

Under IoT we understand the local sensors installed or to be installed in the areas of the Open Labs.

Information on existing sensors

We would like to receive the following information from the Open Lab representatives on the locally installed sensors:

- Which sensors are already installed in the Open labs (e.g. temperature, water level, humidity)? Could you please share with us some information about (e.g. datasheet and/or other documentation)?
- Which data are measured by these sensors and in which units of measure?
- How are these data collected today?
- Where and how are the sensor data stored (e.g. in local or remote servers, in local data loggers, in data bases, type of data base e.g. MySQL, Access, Excel files, etc.)? Is this a proprietary or open system? Could you please share with us some information/documentation about?
- Describe the type of file and the file structure containing the sensor data
- Describe your actual data retrieval process

Are there any available data derived from such sensors? If yes, are such data already described on the Data Mapping Form ?

Information on future sensors

Regarding the sensors under consideration to be installed during Shelter project we would like to receive the following information from the Open lab representatives:

- Which sensors are planned to be installed (preliminary plans were included in the proposal for the different Open Labs)? Could you please share with us some information about (e.g. datasheet and/or other documentation)?
- What parameters will be measured
- What type of existing monitoring network is installed today (wifi, Bluetooth, wired; communication protocols like ZigBee, Modbus, MQTT, etc.; local data storage or via gateway to remote server)
- Will the new sensors have to be integrated in the existing monitoring system(s) or will a new monitoring network be built? Is this a proprietary or open system? Could you please share with us some information/documentation about?

If possible, can you describe the type of file and the file structure containing the sensor data?

10.5 Annex V: Detailed information on IoT sensors and the related data collected through the dedicated survey (see Annex IV)

Baixa-Limia Serra Open Lab:

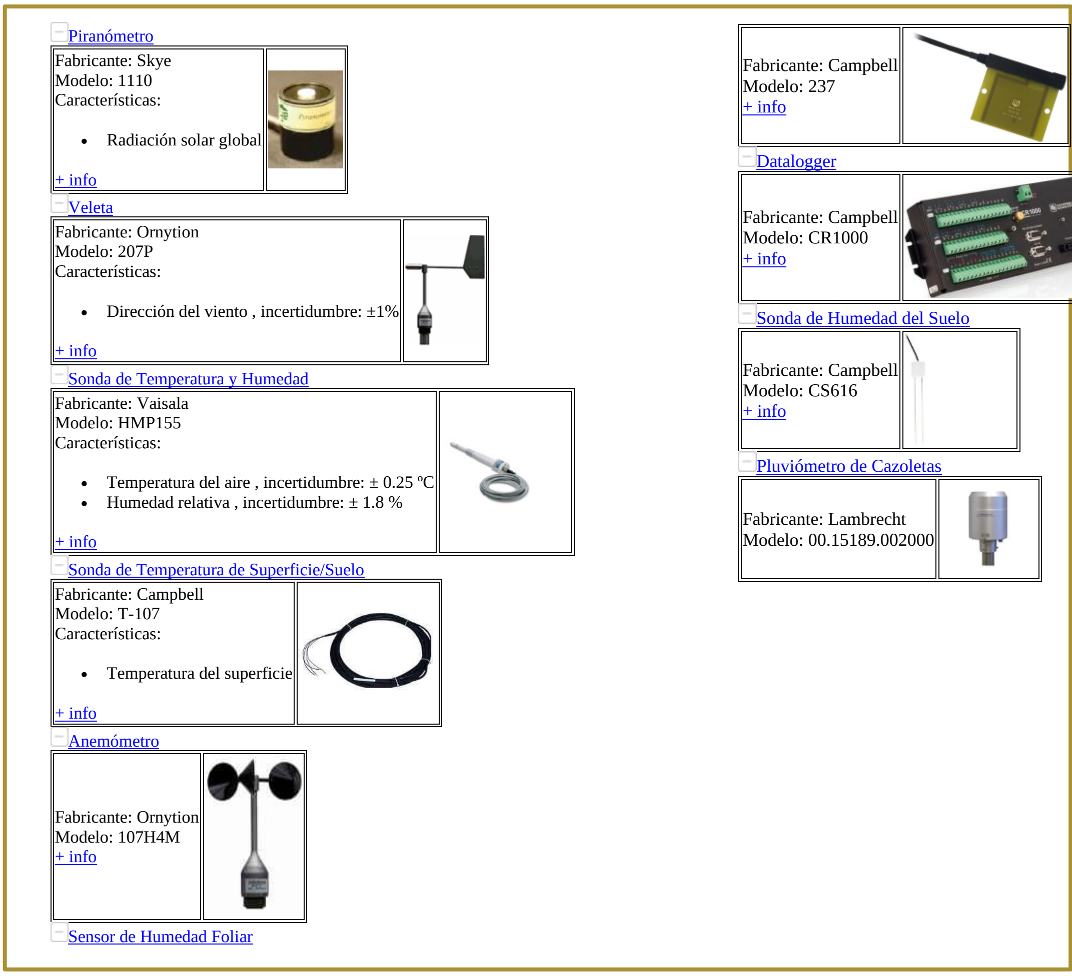


Figure A- 5 - Xunta de Galicia IoT sensors: Calvos de Randìn station (OU)

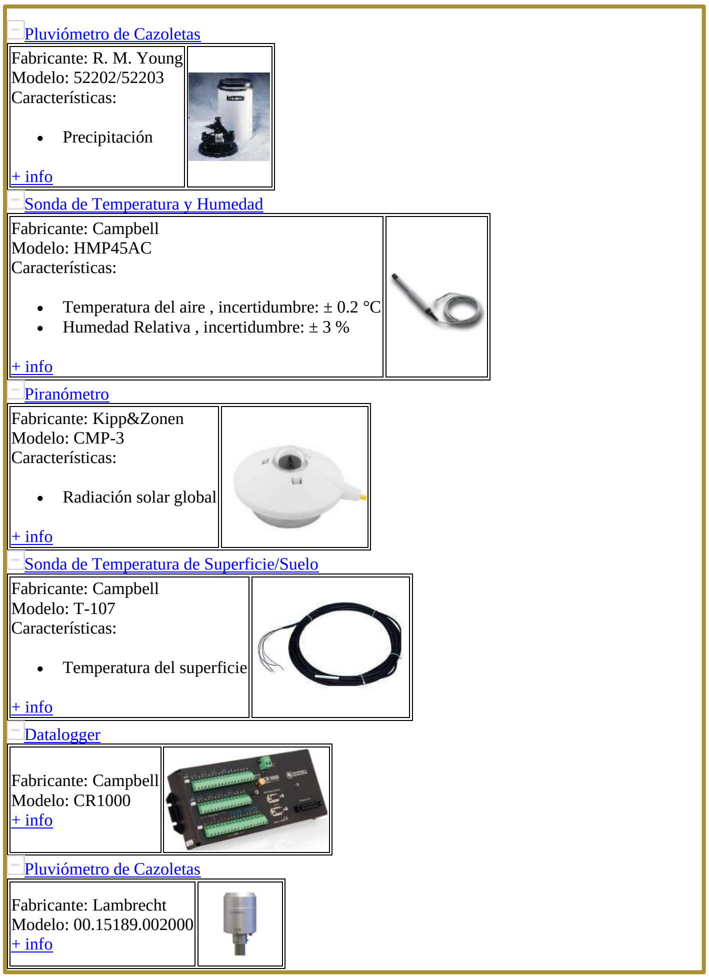


Figure A- 6 – Xunta de Galicia IoT sensors: Entrimo station (OU)

Pluviómetro de Cazoletas

Fabricante: R. M. Young
Modelo: 52202/52203
Características:

- Precipitación

+ info

Veleta

Fabricante: Ornytion
Modelo: 207P
Características:

- Dirección del viento , incertidumbre: ±1%

+ info

Sonda de Presión

Fabricante: Vaisala
Modelo: PTB110
Características:

- Presión atmosférica , incertidumbre: ±0.3 hPa at +20 °C

+ info

Sonda de Temperatura y Humedad

Fabricante: Vaisala
Modelo: HMP155
Características:

- Temperatura del aire , incertidumbre: ± 0.25 ºC
- Humedad relativa , incertidumbre: ± 1.8 %

+ info

Piranómetro

Fabricante: Kipp&Zonen
Modelo: CMP-3
Características:

- Radiación solar global

+ info

Sonda de Temperatura de Superficie/Suelo

Figure A- 7 – Xunta de Galicia IoT sensors: Calvos de Randìn station (OU)

XUNTA DE GALICIA	
Which sensors are already installed in the Open labs (e.g. temperature, water level, humidity)?	4 Metereeological Stations (see sensors in the doc file) • Council of Entrimo • Council of Muiños • Council of Lobios • Council of Calvos de Randín (next to the park, not inside)
Could you please share with us some information about (e.g. datasheet and/or other documentation)?	See example
Which data are measured by these sensors and in which units of measure	Check DATA sheet to see which parameters are measured for each station, and which data are measured 10 min, daily and monthly
How are these data collected today?	CSV
Where and how are the sensor data stored (e.g. in local or remote servers, in local data loggers, in data bases, type of data base e.g. MySQL, Access, Excel files,..)? Is this a proprietary or open system? Could you please share with us some information/documentation about?	The data enters within local data loggers, then they are validated and stored in a data base. This data base is SQG Server, a propietary system. Howeveer the final data is available online and can be downloaded as CSV, JSON or pdf format.
Describe the type of file and the file structure containing the sensor data	See example to see how the final data is structured
Describe your actual data retrieval process	
ailable data derived from such sensors? If yes, are such data already described on the Data	Yes. The website shows prediction maps using these data--> There are reports regarding metereeological data as some reports (unstructured data). Ex: 75023 Monthly reports Metereeological data is also considered to generate the daily risk of fire: 75007
	https://www.meteogalicia.gal/modelos/index.action?request_locale=gl
	https://www.meteogalicia.gal/observacion/informesclima/informesIndex.action
	https://mediorural.xunta.gal/es/temas/defensa-monte/irdi
Information on future sensors	Xunta de Galicia will not install new sensors

Table A- 1 – List of meteorological data collected by the sensors described in figures A-5, -6 and -7.

Dordrecht Open Lab:

Collection of information about IoT sensors and the related data
Requested by: RED, POLITO, LINKS

Definitions

Under IoT we understand the local sensors installed or to be installed in the areas of the Open Labs.

Information on existing sensors

We would like to receive the following information from the Open Lab representatives on the locally installed sensors:

- Which sensors are already installed in the Open labs (e.g. temperature, water level, humidity)? Could you please share with us some information about (e.g. datasheet and/or other documentation)?

*As a municipality we do not have many sensors ourselves. The sensors that we do have are mostly small scale weather station and ground water level monitors. These stations are however not in the old city centre, which is the location of the SHELTER case study. Information that is relevant for the SHELTER project and case study, are supplied by sensors that are governed on a regional or most often national scale. The national agency for infrastructure and water managements (Rijkswaterstaat) has a lot of information available via viewing platforms (*https://waterinfo.rws.nl/#!/nav/publiek/*), this data can also be requested for further analyses. I have already supplied the datasets and links in the Data mapping Form.*

There are further databases with the locations of monumnets (data on municipal and national level) and databases detailing all the waterways, structures and boundaries in relation to the water management system. I can share these links but they area databases, not sensors.

- Which data are measured by these sensors and in which units of measure

The data on the rijkswaterstaat portal includes but is not limited to: water height, wave heights, information on waterflow and water quality. There is a large database of historic data, and a map in which the next 48 hours are forecasted. This data goes back to at least 1945

- How are these data collected today?

There are hundreds of sensors in the watersystems and on the north sea managed by Rijkswaterstaat. These sensors send constant data, that is viewable on an online platform. Overall the Netherlands has a lot of open data available (almost 15.000) datasets including all different subjects. These can be found here: https://data.overheid.nl/. *The datasets from Rijkswaterstaat should suffice for the SHELTER work as far as I know.*

- Where and how are the sensor data stored (e.g. in local or remote servers, in local data loggers, in data bases, type of data base e.g. MySQL, Access, Excel files,..)? Is this a proprietary or open system? Could you please share with us some information/documentation about?

As a municipality we also supply our own data sets that are open (all data on meterological sensors are) to national databases. Rijkswaterstaat also have their own database in which data can be viewed and downloaded. They are shared within the WFS and WMS protocols.

- Describe the type of file and the file structure containing the sensor data

For datasets on water a CSV file can be downloaded easily on this website: https://waterinfo.rws.nl/#!/nav/bulkdownload/huidige-selectie/ *. The data can also be viewed online in a viewing platform*

- Describe your actual data retrieval process

See above.

Are there any available data derived from such sensors? If yes, are such data already described on the Data Mapping Form ?

Yes.

Information on future sensors

We do not plan the installation of any future sensors within the old city centre of Dordrecht, in relation to datasets and information that are needed for the project.