D5.2 Low-Level Design

The architecture of the DIRECTED Data Fabric

12 November 2024

Martin Pontius[15], Julia Kraatz[15], Henning Bredel[15], Benedikt Gräler[15], Eike Hinderk Jürrens[15], Stefano Bagli[4], Francesca Renzi[4], Valerio Luzzi[4], Marco Renzi[4], Paolo Mazzoli[4], Sukaina Bharwani[18], Valeria Pancioli[8], Dominik Hedderich[9], Jana Löhrlein[11], Janne Parviainen[18], Lydia Cumiskey[6], Tobias Conradt[2], Martin Drews[3], Jacob Pedersen[7], Sandro Nanni[13], Christopher Genillard[9], Stefan Hochrainer-Stigler[10], Levente Huszti[12], Heiko Apel[14], Shahin Khosh Bin Ghomash[14], Chahan M. Kropf[16], Julian Struck[11], Benjamin Hofbauer[5], Max Steinhausen[1], Tracy Irvine[17]

1. Technische Universität Braunschweig
2. Potsdam-Institut für Klimafolgenforschung e.V
3. Danmarks Tekniske Universitet
4. Gecosistema SRL
5. Research Institute for Sustainability
6. University College Cork – National University of Ireland, Cork
7. Region Hovedstaden
8. Agenzia Regionale Per La Sicurezza Territoriale E La Protezione Civile
9. Genillard & Co GmbH
10. Internationales Institut für angewandte Systemanalyse
11. Erftverband
12. Zala Különleges Mentők és ÖnkéntesTűzoltó Egyesület
13. Agenzia Regionale per la Prevenzione, l'Ambiente e l'Energia dell'Emilia-Romagna
14. Helmholtz Zentrum Potsdam Deutsches Geoforschungszentrum GFZ
15. 52°North Spatial Information Research GmbH
16. Eidgenössische Technische Hochschule Zürich
17. OASIS Hub Limited
18. SEI Oxford Office Limited

Project: 101073978 — DIRECTED — HORIZON-CL3-2021-DRS-01

Funded by the European Union

Report Overview

Project number	101073978
Project acronym	DIRECTED
Project name	Disaster Resilience for Extreme Climate Events providing Interoperable Data, Models, Communication and Governance
Call	HORIZON-CL3-2021-DRS-01
Topic	HORIZON-CL3-2021-DRS-01-02
Type of action	HORIZON Innovation Action
Responsible service	European Research Executive Agency
Project starting date	01/10/2022
Project duration	4 Years
Period covered	1 October 2022–30 September 2026
Reporting period number	RP2
Periodic report date and version	12 November 2024

Funded by the European Union

Document History

Version	Date	Comment
1.0	2024-10-10	Initial version
2.0	2024-11-12	Submission of draft
3.0		Final submission

Funded by
the European Union

Executive Summary

This deliverable builds upon the high-level architecture of the Data Fabric which was laid out in D5.1. This document, D5.2, describes the low-level architecture of the Data Fabric and describes the technical details. It is written in conjunction with D5.3, where data protection and ethical aspects of the Data-Fabric are documented, and D5.4, where the implementation plan and forthcoming documentation are rolled out. Therefore, several cross-references will be found among D5.1, D5.2, D5.3 and D5.4 altogether documenting the preparations necessary to successfully develop, deploy and operate the Data Fabric in DIRECTED.

The described architecture of the Data Fabric will serve as a reference architecture. It starts from top to bottom, with two main components (Figure 8), the authorization part alongside the Spatial Data Infrastructure (SDI). Separating both follows the modular design of the Data Fabric, ensuring a high degree of re-usability. Both parts are built upon proven Open-Source Solutions in their respective fields.

The SDI covers the entire process from data registration and integration, storage and maintenance up to discovery and access. An orchestration service is introduced that can be used to continuously ingest and update data, as well as allows users to define and execute model chains. All application programming interfaces (APIs) are based on open standards and enable an easy integration with data sources and models outside the DIRECTED project.

The authorization part provides means to facilitate data protection, and secure and legit access to internal and external data sources. Here, the data steward adding and managing data and model chains has a central role to define and to manage access. The defined policies are also enforced in a machine-readable way, such that model chains of users can only access the respective data sources. The metadata is enriched to facilitate and track GDPR compliance,

In order to address auditability and to allow scrutiny of results of the Data Fabric, the model runs and incorporated data is tracked and stored as metadata. This allows to re-produce the model execution and to further validate and audit model outputs.

The architecture design presented in this deliverable is focussed on the thematic user stories and technical requirements of the RWLs in DIRECTED. All RWL use-cases have been further detailed and expanded compared to Deliverable D5.1. The selected user stories per RWL address key challenges of interoperability identified within DIRECTED and will therefore directly contribute to an improved interoperability and the project's DRR and CCA goals.

Funded by the European Union

Contents

Index of Figures

Index of Tables

List of Abbreviations

ACRONYM	DEFINITION
AAI	Authentication and Authorization Infrastructure
API	Application Programming Interface
AuthN	Authentication
AuthZ	Authorization
CCA	Climate Change Adaptation
CDS	(Copernicus) Climate Data Store
COG	Cloud optimised GeoTIFF
CRUD	Create Read Update Delete
DMP	Data Management Plan
DRM	Disaster Risk Management
DRR	Disaster Risk Reduction
FAIR	Findable, Accessible, Interoperable, Reusable
GeoTIFF	Spatial Tagged Image File Format
GIS	Geographic Information System
GUI	Graphical User Interface
HTML	Hypertext Markup Language

JSON	JavaScript Object Notation
k8s	Kubernetes
LMF	Loss Modelling Framework
OGC	Open Geospatial Consortium
OTC	Open Telekom Cloud
RWL	Real World Lab
SDI	Spatial Data Infrastructure
STAC	Spatio Temporal Asset Catalog
WP	Work Package
WFS	Web Feature Service
WMS	Web Map Service
ZARR	community project to develop specifications and software for storage of large N-dimensional typed arrays: https://zarr.dev

1 Introduction

The Data Fabric will showcase how an improved technical interoperability supports better CCA and DRM in the RWLs. The concept of a Data Fabric is outlined in Figure 1 and shows different layers of the system. The horizontal layers depict the technical layers. The vertical layers connect the technical ones based on particular use case requirements. Each RWL brings its own use cases, which may have similar requirements but different foci or priorities. Therefore, the technical architecture has to be able to handle the following challenges:

- Different RWL will have different requirements on the data infrastructure
- For all RWL, not all components/orchestrations have the same importance

The main goal of the architectural design is to define architectural template descriptions which are designed to be composable for tailored use cases. The low-level architecture builds upon the high-level architecture design described by D5.1 and the requirements of the RWLs highlighted therein.

In the following, the Data Fabric layers are explained shortly. Each layer will reference relevant subsections of Section 2 to give a deeper insight of the particular implementation. In Section 3 the implementations are explained in the context of the selected RWL use cases.

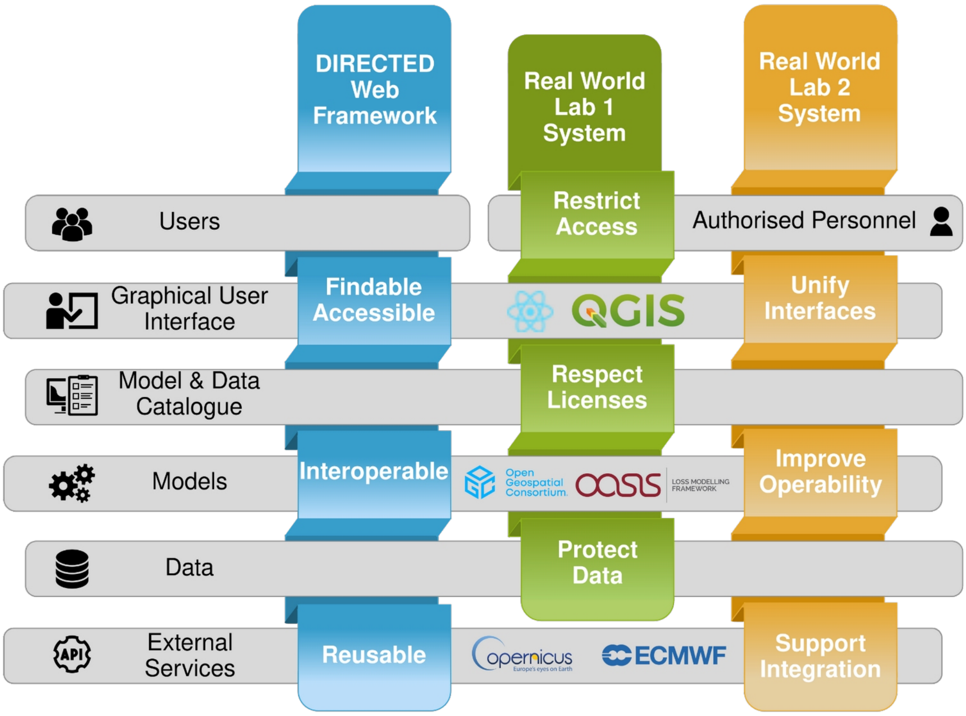

Figure 1: High-Level concept view of the Layers of the DIRECTED Data Fabric.

Users: Access restriction to particular data is a cross-cutting concern. While single services may implement specific authorization rules, the overall approach focuses on a single enforcement point, protecting traffic based on declared rules. User or service based access tokens could also be generated by an external identity provider proxied by the internal Authentication and Authorization Infrastructure (AAI). (see Section 2.7: Data Authorization and Stewardship)

Graphical User Interface: The Graphical User Interfaces (GUI) provides access to the interoperable components and information products. The Data Fabric provides dedicated web applications as well the integration into existing tools and applications (e.g., QGIS) of the RWLs (see Section 2.3: Data Integration and Ingestion and Section 2.8: Applications).

Model & Data Catalogue: Makes metadata about data and process models via standardised metadata descriptions available. Makes them findable via federated catalogue search as a step before service integration. (see Section 2.1: SDI Gateway, Section 2.4: Data Publication and Section 2.5: Data Discovery)

Funded by the European Union

Models: Simplified and composable processing by exposing well-known and necessary process models via interoperable interfaces (i.e., OGC APIs, Oasis Loss Modelling Framework - LMF). (see Section 2.2: Model Interoperability and Process Orchestration)

Data: Connect to different kinds of local data formats, e.g., file based, RWL specific databases. (see Section 2.3: Data Integration and Ingestion)

External Services: Integrate existing data services via API by implementing adapters for needed data sources to leverage flexible data processing and data access via interoperable Web interfaces.

2 Overview of SDI Technical Components in the Data Fabric

This section describes the technical components of the Spatial Data Infrastructure as part of the DIRECTED Data Fabric.

2.1 SDI Gateway

To reach the goals of the DIRECTED project a core requirement is the interoperability of data, models, communication and governance. Interoperability can be facilitated by using open standards (see D2.1). Thus, the Data Fabric will make use of open standards wherever beneficial. A central component to support this is a gateway for the spatial data infrastructure (SDI) based on the OGC API suite of standards (see Deliverable D2.1). With these standards, users will be able to discover, access and process data in an interoperable and clearly defined way. Moreover, integration with other software systems and tools will be facilitated, as many mainstream software components in the geo domain already implement OGC APIs or OGC Web Services.

The gateway serves as a single entry point for the Data Fabric. It forwards requests to internal components which are responsible for processing these and creating a response. An important task of the gateway apart from providing a uniform view on the Data Fabric is the enforcement of defined policies to external traffic via authentication and authorisation components (see Section 2.7). Having a single entry point with access control allows flexibility with the deployment of underlying components which handle the main load when processing user requests. Some advantages are a better scaling, the option to have small instances which are easier to manage, a lightweight handling of traffic and fewer downtimes.

Although the OGC APIs are relatively new (the first OGC API standard was officially approved in 2019) there are already mature implementations. One of them is pygeoapi which is an open-source server implementation written in Python and capable of deploying RESTful OGC API endpoints using OpenAPI, JSON, and HTML (https://pygeoapi.io/). Pygeoapi supports most of the OGC API standards, for some of them being a reference implementation. Reference implementations are officially certified by the OGC after passing a set of predefined tests. Table 6 gives an overview of the implementation status of different OGC APIs by pygeoapi.

Table 1: Overview of the standards implemented by pygeoapi (source: https://docs.pygeoapi.io/en/stable/introduction.html#standards-support).

Standard	Support
OGC API - Features	Reference Implementation
OGC API - Coverages	Implementing
OGC API - Maps	Implementing
OGC API - Tiles	Reference Implementation
OGC API - Processes	Compliant
OGC API - Records	Implementing
OGC API - Environmental Data Retrieval	Reference Implementation
SpatioTemporal Asset Catalog	Implementing

Figure 2 shows the main components of pygeoapi. At its centre is the core API, which provides the functionality to interact with spatial data. It has to be configured by the administrator for the specific deployment, including the registration of datasets and processes. An Open API document is generated automatically by the OpenAPI component. On top of the core API sits a web framework, which is responsible for serving the application over the web. With Flask, Starlette and Django different options are available. An important aspect of pygeoapi's architecture is the flexible plugin mechanism. A number of built-in data provider plugins exist which support storage and publishing of a wide range of data types and formats served from different backends as databases, file systems and services. New provider plugins can be integrated if necessary. Before users can access data via pygeoapi the data have to be ingested into a suitable storage within the Data Fabric (see Section 2.3 and 2.4). Alternatively, pygeoapi can integrate data in a federated way by serving as a proxy for external data sources without storing a copy of the data (see Section 2.3). For example, it offers built-in providers for forwarding requests from/to a Web Map Service (WMS) or Web Feature Service (WFS), or even Sensor Things API (STA). Data from external providers which support only specific, potentially non-standardized access mechanisms could be imported using dedicated processes via the OGC API Processes standard or proxied through other suitable OGC APIs. Processes will also be used for executing the models provided by the different project partners (see Section 2.2). The Oasis Loss Modelling Framework (LMF) will also be wrapped into the OGC patterns, in order to benefit from existing resources in the insurance sector.

By providing a streamlined, standards-based framework for the models and data, users themselves do not have to interact with a multitude of services within the Data Fabric. Specific user interfaces and applications will further lower the barrier of model and data access (see Section 2.8) and enable users to find solutions for their problems at hand (see Chapter 3). This ensures that the user experience is not cumbersome and difficult to understand, which could be a serious barrier to dissemination; not only to the RWLs, but also to the wider communities the RWLs are part of. Thus, using pygeoapi as a proxy within the DIRECTED project offers standardisation through its backing in OGC, and provides a simplified framework for accessing model outputs for specific user requirements and data for model needs.

pygeoapi C4 Component diagram

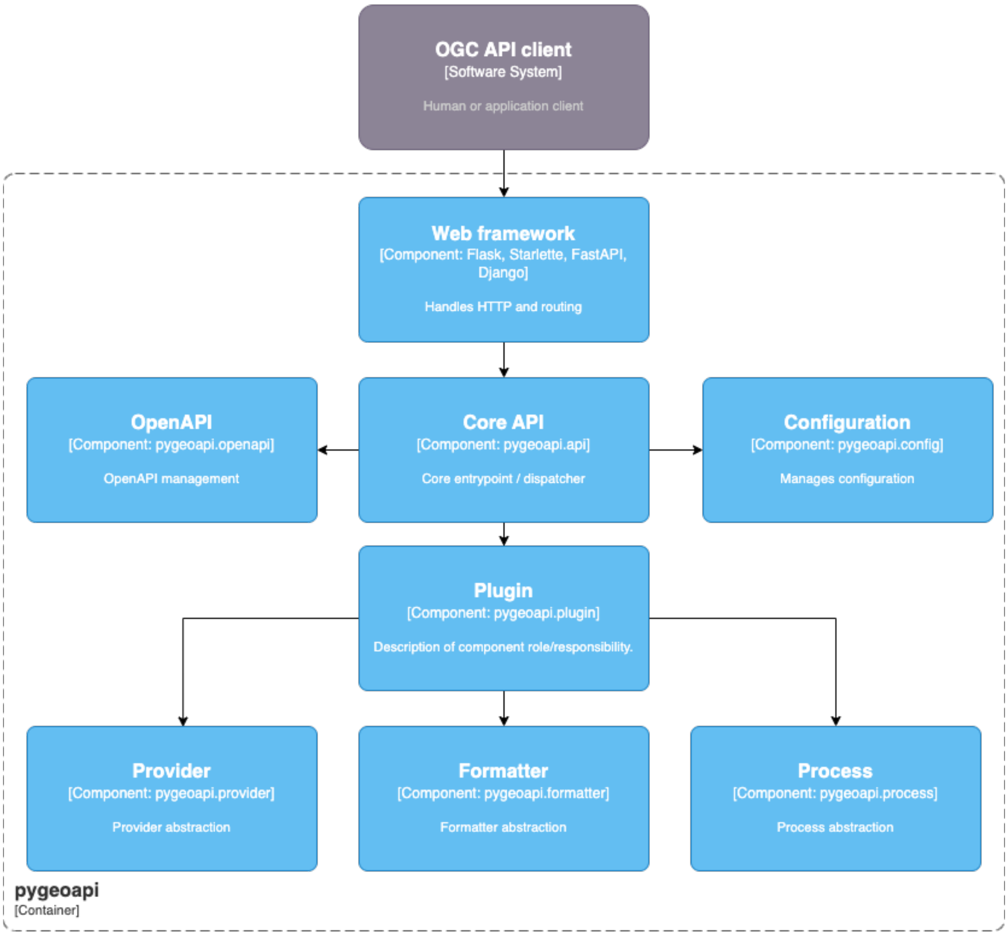

Figure 2: Component diagram of pygeoapi (source: https://docs.pygeoapi.io/en/stable/how-pygeoapi-works.html).

2.2 Model Interoperability and Process Orchestration

As described in Deliverable D5.1 the models brought into the DIRECTED project by different partners of the consortium play a central role in achieving the goals of disaster risk management and climate change adaptation. In this section, we describe the technical aspects of achieving model interoperability and explain the relevant components of the Data Fabric.

When discussing model interoperability on a technical level, mainly two aspects have to be considered:

1. Data and metadata standards (formats/encodings)
2. Interface standards

When using data and metadata standards (see Deliverable D2.1 for a detailed list of options) it is ensured that outputs of a model can be easily used in other software components. Moreover, different data sources, e.g. from data providers or outputs from other models, can be used as input without major efforts in understanding the data and transforming or preprocessing them. Using metadata standards helps understanding and finding data.

At this stage, models still have to be executed manually by someone who understands the tool to a sufficient degree. Moreover, a suitable runtime environment and the necessary hardware has to be available. This makes executing models and especially chaining different models cumbersome and time-consuming. Thus, apart from using data and metadata standards, it is recommended to use interface standards in addition. Candidates for interface standards (OWS, OGC APIs) have previously been described in Deliverable D2.1.

Figure 3 gives an overview of the planned architecture to support model interoperability based on the OGC API Processes standard. A central component is pygeoapi (see Section 2.1) which implements - among others - the OGC API Processes standard. Users (and applications) can make HTTP requests to pygeoapi via the standardised API to run models, check the status of the execution and get the results. Pygeoapi parses the requests and delegates them to a process orchestrator which is responsible for actually executing the processes, considering the specific software and hardware needs. In order to facilitate the orchestration, the models should be wrapped into containerized applications, e.g. via Docker or Kubernetes (K8s).

For the process orchestrator component, several options are available:

- Workflow orchestration frameworks, e.g.
  - prefect (https://www.prefect.io/)
  - Argo Workflows (https://argoproj.github.io/workflows/)
  - Apache Airflow (https://airflow.apache.org/)
- Custom (Python) applications using Kubernetes
  - k8s jobs are created via the k8s API which will execute one of the containerized applications. Communication of inputs/outputs is handled via cloud object storage.
  - pygeoapi deployments are created via the k8s API and subsequently processes are started via OGC API Processes requests. Communication of inputs/outputs is handled via OGC API Processes.

The favoured solution is to use the prefect framework. It provides out-of-the-box features like monitoring, logging, scheduling and a UI. Apart from this, there is already a prefect job manager implementation for pygeoapi (https://github.com/geobeyond/pygeoapi-prefect).

Data ingestion might make use of the same processing infrastructure as well (see Section 2.3 for details).

Funded by the European Union

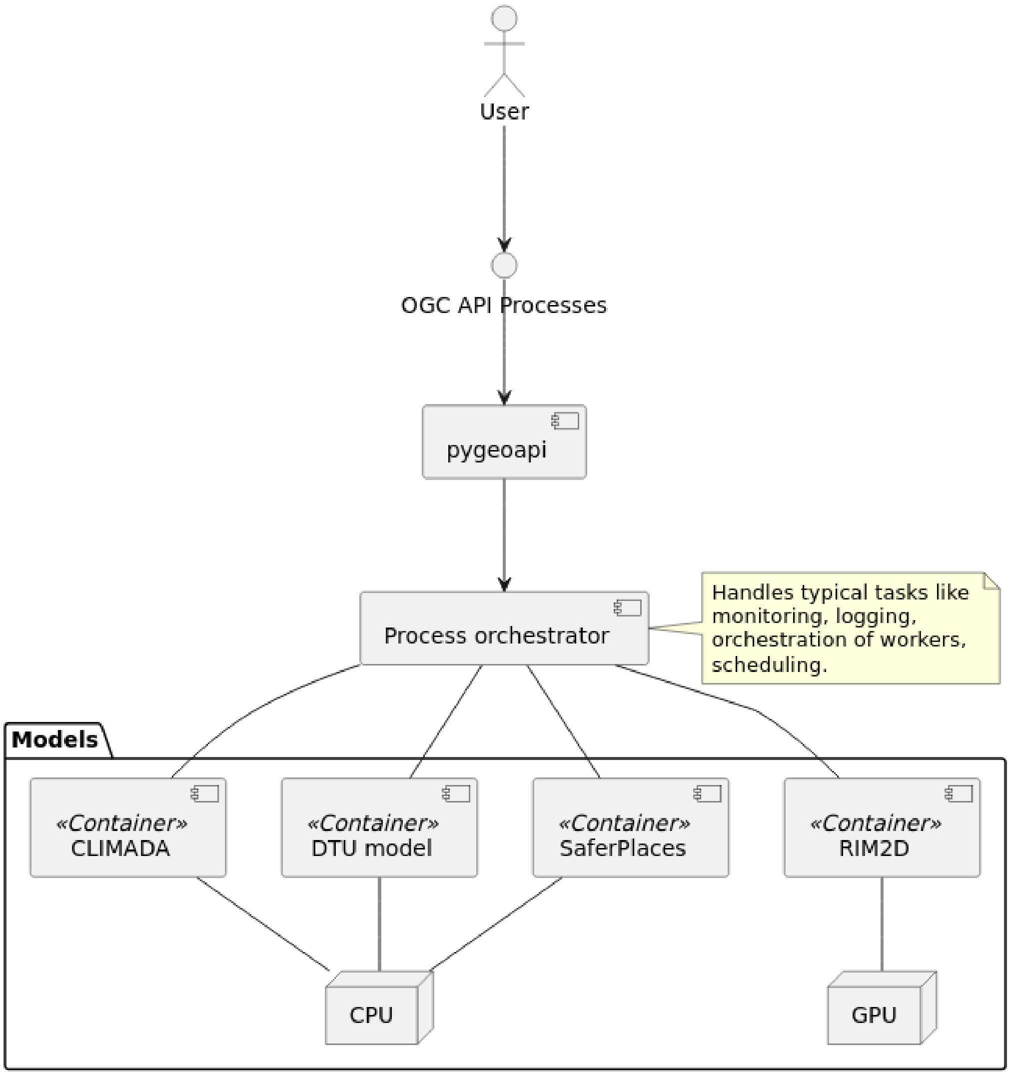

Figure 3: Overview of the architecture supporting model interoperability based on the OGC API Processes standard. The process orchestrator picks the right model based on the request and facilitates its execution, ensuring the availability of the desired hardware.

2.3 Data Integration and Ingestion

Within the Data Fabric, data ingestion and data integration are two different approaches to providing data. Data integration focuses on the integration of external data, which is made discoverable and accessible in the same way as data that is directly available within the Data Fabric. Data ingestion can be seen both as manual work to upload data into the system and as an automated process for continuous data insertion. Manual data ingestion is mainly conducted by a data steward or data administrator by first making the data available in the system and to add metadata about that data.

Certain tools and interfaces are provisioned by the Data Fabric to leverage data integration and ingestion. They are used in intermediary components to seamlessly integrate from various external sources (see Figure 5). Internally, the same tools can also be used to realise data ingestion.

In general, it is advisable to use interoperable service interfaces to make data available, even if individual, non-standardised workflows may remain possible, but stay the exception and are kept internal. The manual provisioning of data is depicted in Figure 6 involving a data steward who publishes local data by uploading it, describing its metadata and registering it at a data discovery service.

Automated data ingestion is outlined in Figure 4 and describes an automated process which has to consider three key steps:

1. External data access: The system connects to and retrieves data from various sources. Access to external data can either be direct or by re-using existing adapter tools provided by the Data Fabric.
2. Data processing: Ingested data undergoes processing, which may include conversion, transformation, merging, and updating of local data and caches.
3. Metadata resolution: The ingestion process automatically extracts metadata from the ingested data. The process shall create and update catalogue records to maintain an up-to-date data inventory.

Concrete architecture and implementation of the pipeline is dependent on things:

- Data source and data format
- Necessary (pre-)processing
- Targeted data storage and format

Therefore, the actual ingestion pipeline can only be described in general here, but can be discussed in more detail on a use case (see Section 3).

An automated data ingestion process can be initiated through multiple mechanisms:

- External API: The process can be triggered via an external API call, providing flexibility for integration with other systems.
- Scheduled Jobs: Internal processes may be triggered by external schedulers, such as local Cron jobs or Kubernetes jobs in cloud environments.

Both approaches can be combined to re-use existing strategies and the given execution environment to trigger and run processes. By wrapping the ingestion process as an OGC process, it can become executable from external systems via process orchestrator (Section 2.2). This would allow the system to “eat its own dogfood” by utilising its own standardised interfaces. Data sources to be used in DIRECTED have been listed in Deliverables D5.1 and D2.1. Selected examples will follow in the subsequent sections.

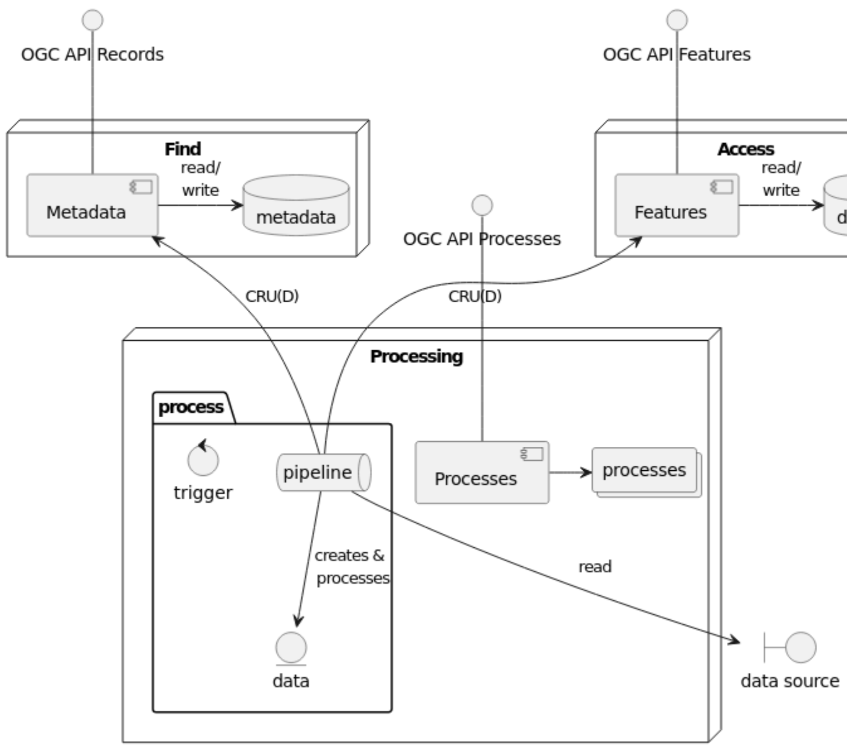

Figure 4: Continuous process of data ingestion. The lower box wraps the process to import external data sources. The pipeline can be triggered manually or automatically (e.g., based on a regular timing). During the integration, the metadata is generated/updated and the data is registered in the data catalogue and made available.

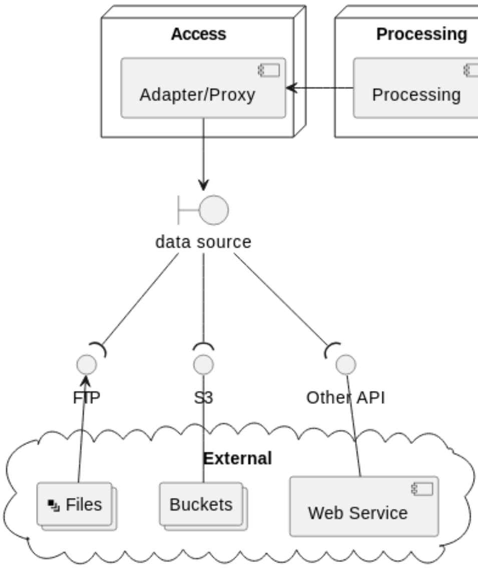

Figure 5: Access of external data via adapter tools. Different dedicated adaptors link to corresponding formats and interfaces for data access.

2.4 Data Publication

As discussed in Section 2.3, data can be ingested and integrated into the Data Fabric using different mechanisms. As a result of data ingestion, the data reside in data storages which can be very heterogeneous and depend - amongst others - on data types, formats, size, intended use and security level. This includes Cloud Object Storage (e.g. Amazon S3) which are regularly used for cloud-optimised raster formats like Cloud Optimized GeoTIFF (COG) or Zarr, databases, files in traditional file systems or services encapsulating and providing data like Elasticsearch instances. At this point, data is only accessible within the system itself. Before external users or applications can use the data, they have to be published in a clear process which includes not only technical but also organisational aspects like access permissions and licences. Figure 6 shows such a process including key actors and components. A data steward publishes data to a store within the Data Fabric. Metadata is derived either automatically or manually by the data steward. It might also be possible to derive parts of the metadata automatically, while other parts need a manual review. In either case the data steward tests and verifies the metadata, which is then stored in a dedicated metadata store, optimally in a standardised format like ISO 19139. Part of this qualified data and metadata publication is the formulation of access permissions and data licences to be compliant with GDPR or other legal requirements. In case data should be published which indicate a high risk to the rights and freedoms of natural persons, a data protection impact assessment (DPIA) has to be part of this process (see D5.3). Once data is stored, the service responsible for publishing the data is configured with the SDI gateway, and metadata is prepared and validated, the actual publication can be concluded.

Users can now either access data directly (in case data location is known to the user beforehand) or the data appears in the result list of some query in a data catalogue where the metadata has been published. Access to data and metadata is possible using standardised APIs (see Section 2.1). It does not matter whether the data is made available via upload, integrated via a data adapter or only contains links to external sources. Depending on the type of data different interoperable APIs might be used:

- Vector data (points, lines, polygons):
    - OGC API Features, OGC API Environmental Data Retrieval
- Raster data
    - OGC API Coverages, OGC API Environmental Data Retrieval
- Sensor time-series data
    - OGC Connected System API
- Maps
    - OGC API Maps, OGC API Tiles
- Metadata

- OGC API Records, STAC

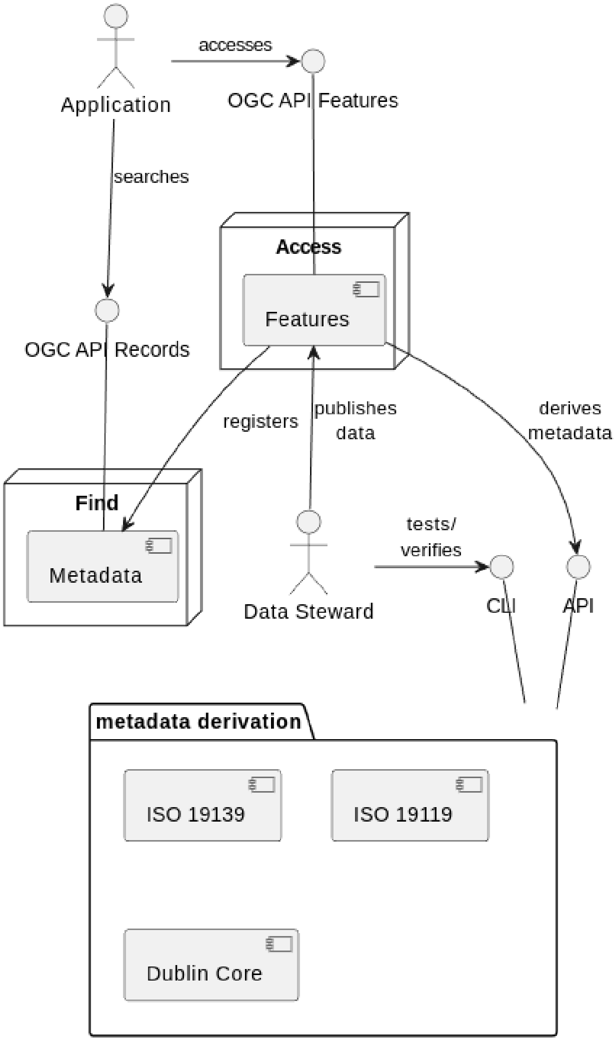

Figure 6: Data publication and access. Data can be found through a dedicated interface (proxied through a web front end). The access of the data (here as an example for type feature) is handled separately. The Data Steward adds and manages data in the Data Fabric.

2.5 Data Discovery

Data from the Data Fabric can only be fully used if it can be easily discovered and integrated. Various types of resources are available within the Data Fabric:

- Data with thematic, spatial and temporal dimensions
- Services that enable access to data and processes
- (Process and model descriptions)

In addition, users can run processes that generate new data based on selected processes and some input data. This output data should also be discoverable, so appropriate metadata descriptions need to be attached to it, just as part of the overall process. A user may not be aware of all resources, so a discovery service is the first address to find out what the Data Fabric has to offer.

The OGC API Web standard family provides different metadata services which allow discovery. Currently, the OGC API for Records is a draft but builds upon common OGC API patterns to provide metadata records under a collection. We will establish a hybrid approach in the Data Fabric that combines the OGC API Records for dynamically updated data resources and the Spatio Temporal Asset Catalog (STAC) for external data, as e.g. the Copernicus Climate Data Store (CDS). The search interface for the user will hide this complexity and no prior knowledge of how or where to find the specific data source will be required.

In order to support the user by data discovery, the data integration pipelines can update the metadata and calculate facets and statistics that help to describe and to identify data resources (see Figure 7).

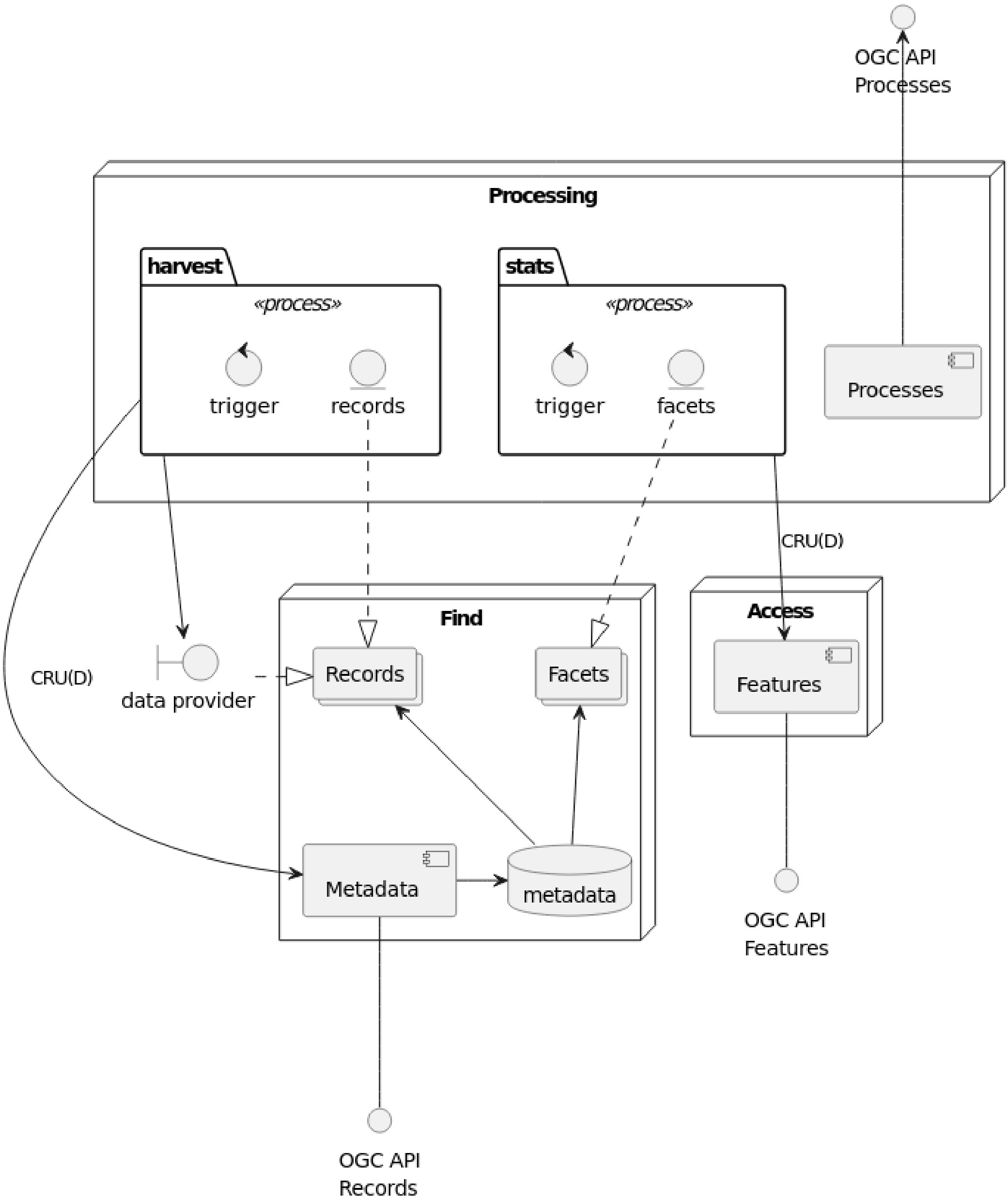

Figure 7: Overview of the interplay between ingestion pipelines and discovery of data sets. Metadata is created or updated during the integration process. The calculation of facets and statistics support the discovery process.

2.6 Taxonomy

The connection between data, models, and metadata is a crucial aspect in keeping consistency with the work in the Data Fabric, as well as the findability, and reusability, of the project's resources. Thus, data within the consortium has been closely documented and monitored in accordance with a metadata naming structure to support the Connectivity Hub (developed by SEI), an open-source taxonomy based in-part from the outcomes of the Data Fabric (see T2.1).

[To be expanded in final version]

2.7 Data Authorization, Authentication and Stewardship

Managing access and usage restrictions of data and services is an important cross-cutting concern of the Data Fabric. The reasons to restrict access are manifold and reach from not yet published data, to sensitive, i.e. safety relevant, data. To avoid the impact on unintended use or even misuse of available data, authentication and authorization components provide technical barriers to protect such data, services, and model outputs.

Storing personal data is an integral part of authentication and authorization as it relies on permissions assigned to a user (directly or indirectly). However, the relevant information of a user will be kept at a minimum (i.e., name, email-address). This data is not a high risk to the rights and freedom of natural persons. Therefore, conducting a data protection impact assessment as described in D5.3 is not necessary.

During the development of the Data Fabric the following aspects are considered:

- The accessing application acting on behalf of some user
- Well-known authentication protocols and standards
- Definition and enforcement of accessing restriction via policies

Revisiting the architecture overview from D5.1 (see Figure 8), the authentication and authorization component runs separately from the individual services proxied by the SDI gateway (see Section 2.1). Together with the gateway, the configuration of policy enforcement becomes centralised and easier to be managed by administrators and data stewards. Users, roles and groups can be created and reused over a set of policies and services. Once a user is authenticated (via OAuth2) and a valid access token has been provided, the following access to the actual data collections are granted (or not) per-request base.

Policies are defined by the data stewards, who are dedicated staff members that can add data sources and model configurations and control the access rights to models and data sets. Policies may be as simple as doing request parsing, but may get more complex when response filtering is required. Such filtering can be realised either with a simple request rewrite or by parsing and filtering the response. In case request rewriting is not possible (setting appropriate query filters transparent to the user) it may become a performance bottleneck quickly. Response filtering would then be better implemented by the individual service. However, serving particular subsets of the requested data was no requirement of any RWL so the centralised enforcement has been prioritised.

The definition of policies is part of the publication process (see Section 2.4). The result is a formal description document (policy) which can be interpreted by the policy enforcement component. The Data Fabric can support data stewards by providing predefined policy templates, for example for a dedicated audience:

- Public
- Owner only
- Members of a RWL
- All DIRECTED members

Within DIRECTED the authorization component aims at further tooling to make writing access policies as easy as possible.

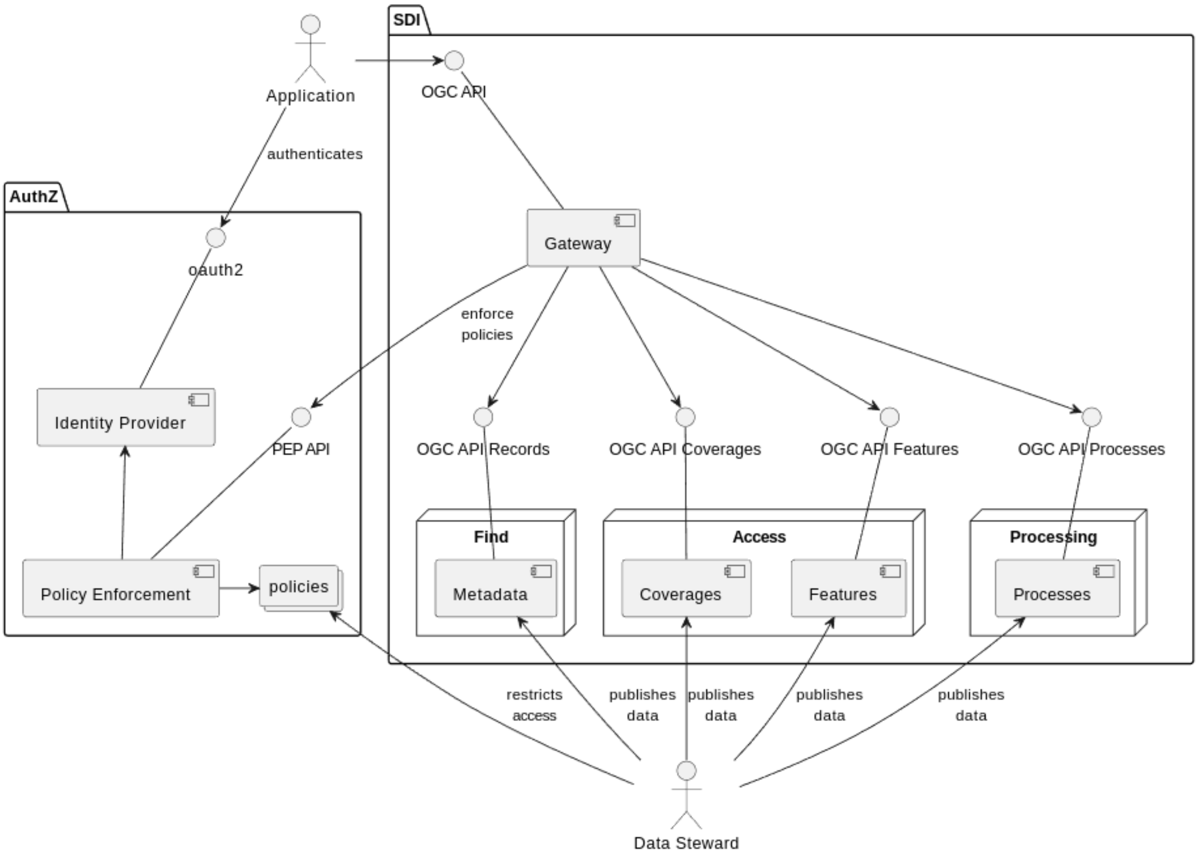

Figure 8: Technical Overview of the Data Fabric with its two core parts: The AuthZ component (left) that manages access by enforcing policies. The SDI components (right) that allow to find, access and process data relevant in the DIRECTED Data Fabric. The Data Steward has a central role in both component sets.

2.8 Applications

As described in the previous sections, the Data Fabric will offer access to a multitude of data and models. In order to enable their practical use by heterogeneous groups of stakeholders, it is necessary to provide suitable applications allowing them to view and interact with the data and models. While data visualisation and user interaction will be heavily influenced by and designed for the RWLs as further specified in Section 3, in this section the different avenues in which the RWLs could potentially choose according to their stakeholders needs will be discussed. The following list gives a general overview of options, which will be explained in more detail afterwards (the numbers are not meant as ranking):

1. GIS layers which can be integrated into existing software solutions
2. Jupyter Notebooks
3. Customised web applications including, e.g., data dashboards or maps
4. Story maps
5. Existing solutions like SaferPlaces

(1) Geographic information systems (GIS) are software applications with a user interface - either a desktop or web-based user interface - capable of executing geospatial processes as well as visualising data as layers on top of a basemap. Stakeholder groups have noted that GIS systems are oftentimes too complex for them to use, and not easily used in times of emergency. As GIS needs to be run with user knowledge, this is understandable; however, storing GIS layers for future reference could still be an advantage within the DIRECTED project. The layers could then flexibly be integrated into widely used GIS software like QGIS or ArcGIS or into systems already developed by stakeholders of the RWL, e.g., the HORA platform in the Danube region RWL. Thus, one option for data visualisation is to have GIS layers available via OGC standards served by, e.g., pygeoapi or GeoServer allowing further analysis or visualisation of project results outside the Data Fabric.

(2) Jupyter Notebooks are a way to run processes step-by-step in order to achieve some end goal. In addition, it is possible to mix in documentation or instructions in a visually appealing way or even to add widgets for user interaction or dynamic display of data. In the context of data visualisation, this could mean that users could go through the code of CLIMADA for simulating high levels of precipitation and storm surge in Roskilde Fjord, and viewing the flood map at the end of the notebook, displayed, potentially, as a layer on a Leaflet map. This would also provide a tool to contribute to the e-learning components of the DIRECTED project. However, Jupyter-Notebooks only target the rather skilled modeller willing to assess a pre-scripted analysis.

(3) Customised web applications could be built, tailored to the needs of the RWLs. These could include maps or data dashboards. Data dashboards can be standalone, or they could be linked to a map by providing detailed information about either a) a set region, point, or

boundary in the map, or b) information based on parameters that a user sets. Depending on the user requirements, data can also be displayed at any timescale which is feasible from the data source. For example, ECMWF offers seasonal forecasts available at a daily and sub-daily (6h) update frequency. If a variable from this dataset, such as precipitation, is specified by the end-user that they would like to visualise, a dashboard would be a beneficial way to display the values of precipitation at points or locations the user clicks in a web map. Web applications can be built using a wide range of programming languages, libraries and frameworks (e.g. React, Angular, Open Pioneer Trails, Streamlit, PyShiny). Some general useful components independent of the RWLs are:

- a catalogue for data and processes: this will be a way for users to find data and processes helpful for their specific use cases. Options to filter by, e.g., format, topic or spatial and temporal extent and resolution should be included among other features which could be requested by the RWLs.
- a simple user interface for running models/model chains: models provided as processes via the OGC API Processes standard have to be triggered using an HTTP request. This could be a hurdle for non-expert users. A user interface which hides these technical parts and allows viewing and running models via clicking a button will therefore be helpful to open models to a broader user group.
- a simple user interface for data stewards to manage policies: defining policies to control access to data and processes needs some expert knowledge. Depending on who will take the role of a data steward, this could make it challenging. A user interface could support data stewards, here also providing templates for basic cases.

Discussions, especially with the Emilia Romagna RWL, have further shown that there is a need for a web-based application supporting the coordination of volunteers during a disaster event.

(4) Story Maps are another type of visualisation that is a blend between a presentation and dynamic geographical information. Instead of showing a static map in a slide, for example, Story Maps enable the integration of web maps alongside static text or pictures to convey spatial information in more detail. This data visualisation could be valuable for dissemination or educational purposes, as it could enable the broader public to see a story about the data and Data Fabric elements as a whole without needing to have someone presenting it to them. Additionally, Story Maps could be used toward the end of the project within the consortium, to showcase the Data Fabric to the RWL stakeholders. Story Maps could be developed by software engineers similar to the customised web applications. However, there are also tools which allow you to create Story Maps with a user interface and supporting features like drag-and-drop (e.g. https://mapstories.de/en). This would enable the direct use by non-technical stakeholders and data stewards could add further Story Maps as needs arise.

(5) Another option is to set up existing solutions like SaferPlaces for the RWLs. SaferPlaces does not only provide models to simulate, e.g., flood events but also includes a sophisticated user interface which allows users to view and upload datasets and to trigger model runs. Integrating existing tools is i.e., advantageous for trained expert users who need to have full control over the modelling chain.

The development of applications depends on the actual needs of the RWL stakeholders and will be done in a co-creation process. Section 3 will present more details on the user needs. Mock-ups of visualisations will be presented in Deliverable D5.4, Implementation Documentation of the Data Fabric.

Funded by the European Union

2.9 Deployment and Runtime

The project deploys its application infrastructure within an Open Telekom Cloud-managed (OTC) Kubernetes (K8s) cluster, specifically within a dedicated and isolated namespace to ensure security and management efficiency. The cluster is running in the German datacenter of the OTC, because of GDPR requirements (see D5.3 "Data Protection & Impact Assessment" for more details). Access to the cluster resources is governed through the Kubernetes API server combined with OTC's Identity and Access Management (IAM) user controls, restricting user permissions strictly to namespace maintenance activities. To protect sensitive data, storage classes employ encryption techniques to prevent data leakage. To add another layer of protection, only the necessary applications have exposure to the internet, while all other components are confined within the project's cluster and namespace. Additionally, network policies are established to manage access to pods that expose ports, and the default OTC firewall configuration restricts incoming connections solely to ports 80 and 443. Traffic is routed to application pods through a load balancer for optimal control and access management. Reliability is ensured through Kubernetes' Horizontal Pod Autoscaler and Node Autoscaler, which automatically adjust resources to maintain uptime as demand fluctuates.

The codebase of each application is hosted on GitHub, where repository access is controlled by defined user groups and the visibility settings that correspond with project requirements. Applications are containerized and published to Docker Hub, facilitating efficient and consistent deployment. The infrastructure relies on Infrastructure as Code (IaC), with dedicated repositories hosted on GitHub separate from application code. This approach reduces visibility, limits access, and centralises control, as all project-specific resources are encapsulated within a single namespace and managed using Kubernetes' Kustomization tool.

Automated pipelines (see Figure 2.9.1) are implemented through GitHub workflows, where tagging the application code initiates builds, image uploads, and security scans via Trivy. Critical build information, such as sensitive configuration values, is stored securely in GitHub repository secrets. Docker Hub and the infrastructure's GitHub repositories are directly integrated with these workflows, allowing updates and deployments to be triggered by commits to the namespace repository or application repository. Trivy is employed twice within the pipeline and during runtime through the k8s Trivy Operator, ensuring continuous security assessment (see D5.3, Section 3.2.2).

Monitoring and alerting are achieved via a Prometheus operator within the same Kubernetes cluster. External endpoint monitoring ensures fast and efficient alerting in case of any downtime of any application within the infrastructure. Alerts based on the monitored metrics are sent through communication channels such as Slack, email, and others.

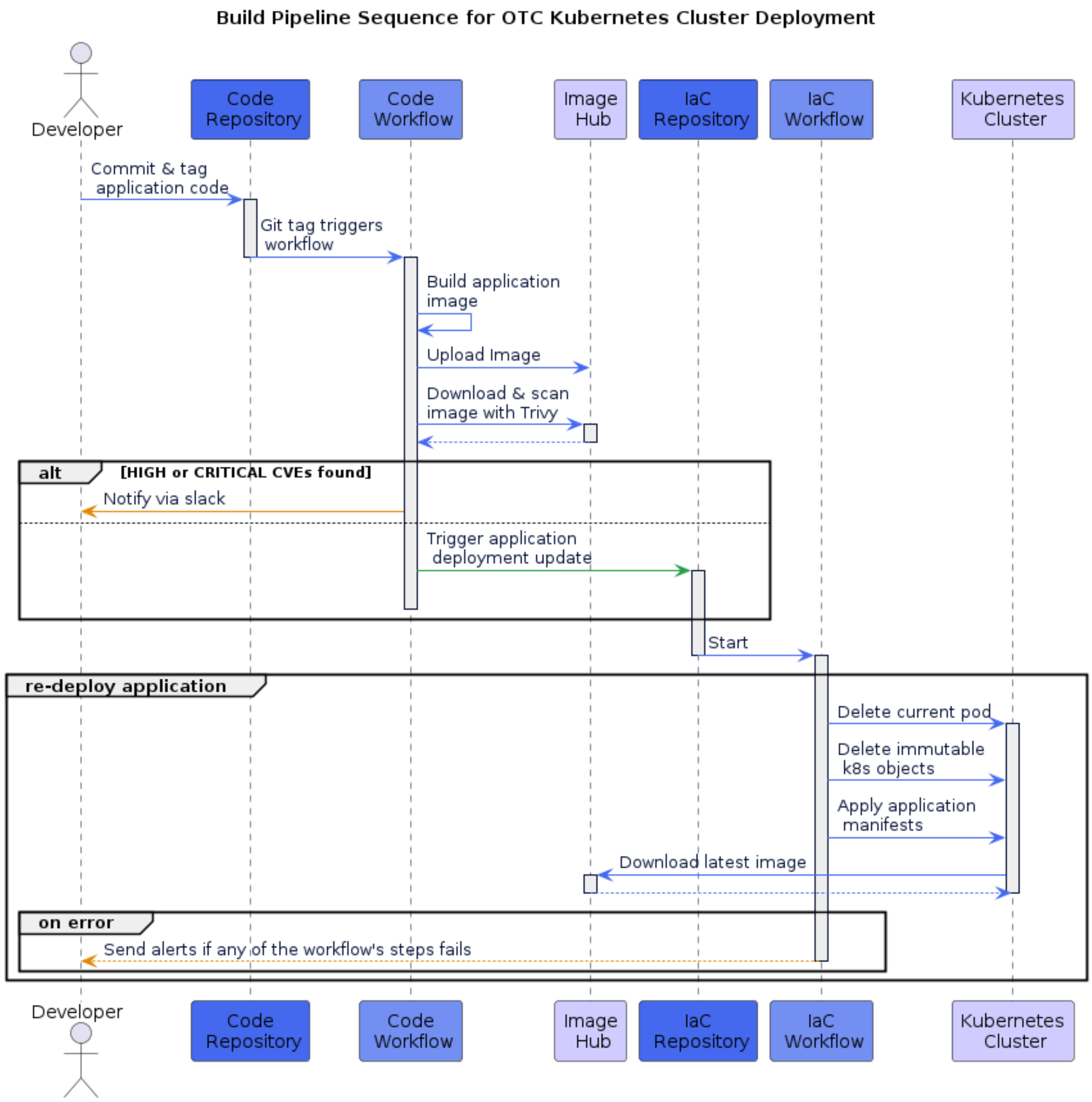

Figure 9: Build Pipeline from code change by a developer until the application in the cluster is updated.

3 Real World Lab Interoperability Use Cases

This section provides an overview how the interoperability use-cases already sketched in Deliverable D5.1 have evolved. The realisation of the user stories will be facilitated based on the technical components described in the Section 2. To ensure meeting the RWL needs, the implementation phase will further incorporate the RWL hosts and stakeholders in feedback loops.

For each RWL, we will provide the personas and user stories that have been identified through collaborative discussions and workshops with RWL hosts, stakeholders and the DIRECTED team. Personas describe a synthetic person which is representative for a user group. This description helps to better understand and detail the requirements in the context of the user. The user stories summarise in compact form the aspects:

- **Who** wants to do/achieve something (role)?
- **What** does this persona want to do (goal)?
- **Why** does this user want it (benefit)?
- **How** does the persona envision to achieve it?

This schema deviates slightly from the classical pattern where “how” is not part of a typical user story. During the discussions with our RWL hosts and stakeholders, the four questions turned out to be more comprehensive.

[Personas and user stories are still undergoing review, prioritisation and selection by the RWLs; the selection will become apparent in the final submission of this deliverable.]

3.1 RWL 1 - Copenhagen Capital Region, Denmark

Personas

The following tables provide structured information on the personas used in this RWL.

Title	**GIS Specialist at Xx**
Name	
Organisation	
Responsibility (in general)	To analyse flood information from DMI in a GIS and inform the municipal authorities in times of potential high precipitation and storm surge

Title	**Climate Modeller in the Roskilde Fjord**
Name	Danish Emergency Management Agency? (DEMA)
Organisation	
Responsibility (in general)	Analysing precipitation data for flood risks and issuing warnings for the municipalities surrounding Roskilde Fjord

Title	**Xx (Risk governance process manager?)**
Name	Ministry of Social Security and Emergency Preparedness? (planned to absorb and oversee risk governance processes, especially in terms of coordinating flood management and coastal protection)
Organisation	
Responsibility (in general)	Architectural design and city planning along the coast of Roskilde Fjord to prevent seawater intrusion

Title	EMS Volunteer
Name	
Organisation	
Responsibility (in general)	Citizen volunteer in Xx who aids in decision-making and response during heavy rainfall events; risk communication with citizens

User Stories

Title	**GIS Flood-Map Layers and Information Sharing among the Municipalities (Xx Governing Body)
Who?	GIS knowledgeable user
What?	As a GIS specialist, I look at precipitation data and flood-maps within QGIS (or ArcGIS) to make inferences about upcoming weather forecasts and potential hazard events. As this information is not readily understandable to my colleagues when I am unavailable, I need a software that is understandable and accessible in a GIS or WebGIS platform for my colleagues for precipitation and storm-surge related data
Why?	This would help in times of my absence, so that my colleagues are able to view the information, and inform neighbouring municipalities during potential hazard events (namely, high-precipitation and storm surge)
How?	I would like to have flood-maps containing data in real-time via DMI; ideally these would also be attached to a WebGIS dashboard so that a discharge hydrograph can be viewed alongside the data and current storm-surge boundaries/physical barriers –simpler the better for UI –learning modules?? try

Title	⭐ **Open-Source Models able to Couple Between Persistent Rainfall and High Water Levels in the Roskilde Fjord (Xx Organization)**
Who?	Danish Emergency Management Agency? (DEMA)
What?	As a Climate Modeller(?), I run (xx models) to analyse rainfall data and water levels in the Roskilde Fjord, to identify potential persistent rainfall events and high water levels in the Roskilde Fjord; I do this to issue warnings to the surrounding municipalities in times of potential hazard events. However, the current models we use are not open-source and do not couple between these two weather phenomena, and I would like to utilise the models in DIRECTED to do so
Why?	This would aid my organisation in the warnings we issue to the municipalities, because we would have a better picture of the severity of the potential hazard event. When these two weather phenomena occur instantaneously, they have had disastrous effects in the past. Thus, having models that couple these events would mean that our warnings are more robust
How?	I would like the modelling chain containing the coupled precipitation and water level data to be run daily, and the outputs to be stored in our local server for historical reference. During particularly hazardous events, I would like to be able to run the DTU Damage-Cost Model on demand, to see the potential impact of the event. We could then issue warnings accordingly

Title	⭐ **Adaptation Measures for Seawater Intrusion (Xx Organization)**
Who?	Ministry of Social Security and Emergency Preparedness? (planned to absorb and oversee risk governance processes, especially in terms of coordinating flood management and coastal protection)
What?	As an Xx, I look at areas in which to place barriers along the coast of Roskilde Fjord to prevent seawater intrusion. However, some of the old infrastructure is in need of updating–some of the barriers are no longer helping, but rather hindering the flow of seawater, causing river flooding. Thus, I would like to use the models in the DIRECTED context to model where old barriers could be removed, and new ones could potentially be placed

Why?	This would help the coastal towns along the Roskilde Fjord with river flooding, and save the cities money with regard to flooding which could be easily prevented
How?	I would like to see flood simulations with and without physical barriers, both in the current state and with barriers taken away. I would like to see these simulation outputs in an interactive Web Map to show to our stakeholders/the governing bodies of the towns surrounding the Fjord.

Title	**Uncertainty predictions (modify with below)**
Who?	EMS Volunteer
What?	As a volunteer for emergency management services in Xx, I respond to warnings sent out by the municipal authorities and respond during hazard events. However, I do not have access to gauging station data which would be helpful in monitoring events/coordinating between other volunteers during heavy rainfall events –unreliable data from DMI; closed data from the police force
Why?	Having readily available gauging station data would be helpful for our planning efforts, so that we are pre-emptively aware of the situation before warnings are issued; additionally, it would help us know where to focus our efforts and divide-up volunteer response –better predictions would mean less wasted hours/efforts (currently happens where events do not occur)
How?	I would like to have real-time data in a Web Application to view from the gauging stations in Xx. I would like to be able to click on these gauging stations and see data from the past 72 hours, as well as forecasted precipitation for the next 24 hours. I would like to be able to view historical highs and lows (i.e. the levels from Storm Bodil) so that I have a reference for high-precipitation. –communications via uncertainty predictions, plethora of values/possible values

Title	**Planning Perspective**

Who?	EU Flooding Directive? (partners with Danish Coastal Authority) –5 municipalities, –and 3 emergency management orgs responsible for: ➔ Action plans ➔ Goals ➔ Visualisations –Danish coastal authority provides data –more detailed maps
What?	
Why?	
How?	How DIRECTED could potentially support their goals/improved technical and communication gaps –***maybe the Data Fabric could be a way to aid the flooding directive with data integration/a common platform they can use

Title	**Standardised Coastal Flood-Risk Warnings**
Who?	EU Flooding Directive? (partners with Danish Coastal Authority)
What?	
Why?	
How?	

Funded by the European Union

Title	Climate Change Scenarios to Communicate Uncertainty in the Lens of Sea Level Rise
Who?	Ministry of Social Security and Emergency Preparedness? (planned to absorb and oversee risk governance processes, especially in terms of coordinating flood management and coastal protection)
What?	The weather and storm forecast models currently used by DMI present challenges when it comes to communicating uncertainty of climate change scenarios; these outputs are not easily understood, and are as precise or accurate as we would like
Why?	Increased accuracy of hazard and forecast warnings would help reduce the currently high costs of mobilising a deploying emergency resource
How?	As the current warning system from DMI issues warnings for the Danish Land and Coastal areas in accordance with the Danish Emergency Management agency 36 hours before a dangerous weather event is expected, having these warnings 72 hours in advance would give more time to mobilise emergency efforts/prevention measures. I would like these measures to be viewable by the public, as they are not currently available to the public on DMI.dk.

Data

[to be completed upon prioritisation and selection of user stories & survey sent out by Martin D.]

Data sources which could be utilised in the Copenhagen Capital Region, specifically targeted toward the Roskilde Fjord, are included in the table below.

Model	Dataset	Variables of Interest	Spatial/Temporal Resolution

SaferPlaces	DMI Historical Sea Levels Observed in the Roskilde Fjord DMI Projected Sea Level Rise Copernicus CMIP6 Climate Projections	DMI Historical: --sealev_dvr --sealev_ln --sea_reg --tw DMI Projected: --water level and storm surge	DMI Historical: unknown spatial res, Denmark, 10min-hourly DMI Projected: unknown spatial res, Denmark, reference period 1981-2010
RIM2D			
SCALGO			
CLIMADA			
DTU Damage-Cost Model			

Components used from the Data Fabric

[to be completed upon prioritisation and selection of user stories]

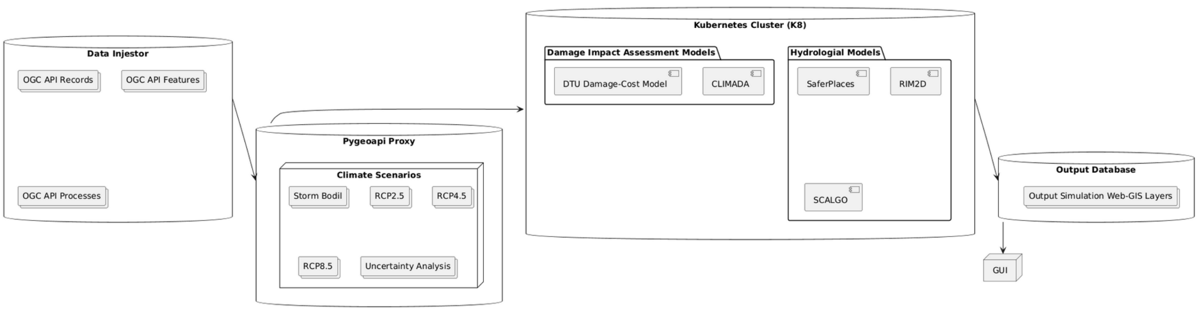

Figure 10: data sources and processing components of RWL 1.

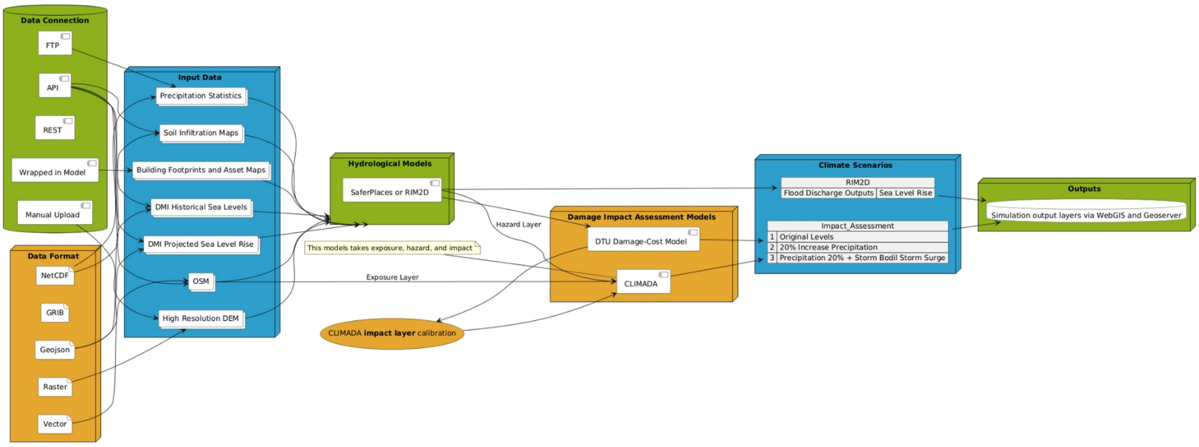

Figure 11: modelling chain for RWL 1.

3.2 RWL 2 - Emilia Romagna Region, Italy

The Emilia-Romagna RWL contribution to developing the Data Fabric builds upon the high-level interoperability architecture outlined in D5.1, specifically focusing on disaster risk management (DRM) and climate change adaptation (CCA), expanding on the technical details of the interoperability framework, emphasising envisaged practical applications in the Emilia-Romagna region. Through the Data Fabric's integration with local tools and systems, monitored real-time data, and models like pygeoapi and cloud-based infrastructures such as AWS, the deliverable outlines how the tools can be leveraged to meet stakeholder requirements.

Stakeholder feedback from workshops and meetings (e.g., March 2023 in Bologna, September 2023 in Ferrara) highlighted the need for improved forecasting capabilities and real-time flood monitoring. The Emilia-Romagna Data Fabric use-case applies these inputs by focusing on improving coordination between forecasting tools, hazard monitoring platforms, and multi-stakeholder decision-making processes (D1.1 Updated_DIRECTED_R…)(D5_1_High_Level_Design_…).

The Real World Lab (RWL) 2 Emilia Romagna is focused on supporting different users in both flood and wildfire risk reduction in two distinct locations: Rimini and Mesola.

Concerning the RWL2 needs for flood risk reduction and adaptation, user needs, personas, and user stories are summarised as follows:

Personas

Title	Regional Civil Protection Officer
Name	Antonio/Christian
Organisation	Civil Protection at regional level
Responsibility (in general)	In case of potential pluvial and coastal flood events, they inform (early warning) and coordinate with municipal Civil Protection Officers for effective engagement and participation in disaster response and preparedness. Plans mitigations such as physical barriers and volunteer coordination.

Title	Municipal Civil Protection Officer
Name	Pietro
Organisation	Civil Protection at municipality level
Responsibility (in general)	Coordination of volunteers for Disaster Risk Reduction during a flood event. Support early warning and emergency operation. Coordinate the design of the Municipal Civil Protection Plan and Climate Change Adaptation.

User Stories

Title	Early Warning and Disaster Risk Management (Municipal Civil Protection Agency)
What?	As a regional civil protection officer, I need a rapid flood modelling tool and

	interoperable web platform that integrates real-time data (both nowcasting sensors, radar and sea/weather forecasts) to produce high resolution data of near-real time and short term forecast of pluvial and coastal flood events.
Why?	This helps in improving disaster preparedness and deploying resources more effectively. For example, designing and localising physical protection barriers or managing the volunteers properly.
How?	The forecasts should be communicated through … maps/time series/traffic lights/calendar widget/... In an existing system, a newly developed web application?

Title	**Flood Resilience and Climate Change Adaptation (Municipal Civil Protection Agency)**
What?	As a municipal civil protection officer, I need a platform capable of mapping flood hazards and damage, considering historical and projected climate data
Why?	This supports my work in adapting the civil protection plans to future climate conditions - Municipal Civil Protection Plan
How?	A cloud web app able to quickly generate flood scenario in a changing climate and changing urban environment

Data

[to be completed upon prioritisation and selection of user stories]

- Key Challenges:

  - The need for real-time (reducing the lag of data from ARPAE), high-resolution flood modeling.

  - Effective communication and governance between civil protection teams and citizens.

  - Climate change adaptation in flood-prone areas.

- Proposed Solutions:

- Data Near Real Time and Forecasting from ARPAE

- Integration of high-resolution models (SaferPlaces, RIM2D) with real-time data.

- Workshops and co-production processes involving citizens and municipalities.

- Simplified Web-GIS applications for easy data access and visualization.

The architecture of the RWL2 Data Fabric shown in the diagrams aims to integrate real-time data, flood modeling, and hazard mapping, addressing user needs for disaster risk management and climate change adaptation.

Components used from the Data Fabric

1. Data Ingestor and Data Pre-Processing:

- Handles real-time observations from HERA radar, ARPAE open data (e.g., river levels, rainfall intensity, sea wave data), and various nowcasting data

- Integrates with ARPAE Forecast Models (e.g., ICON, COSMO2I, COSMO5M) for weather and sea forecasts.

- Data sources include S3 Bucket Static Data (LIDAR, census data, land use, lithology) and external APIs for data from stakeholders and global datasets (e.g., Copernicus CDS).

2. PyGEOAPI Models - processes:

- An API that connects modeling tools like SaferPlaces, RIM2D, CLIMADA, and SVI (social vulnerability index) with real-time and static data.

- Produces high-resolution flood maps and hazard assessments that are accessible through a web platform.

3. K8s Orchestrator:

- Coordinates data flow between the ingesting processes, modelling tools, and the visualisation platform.

- Ensures interoperability and real-time updates of hazard maps and forecasts for users.

4. Visualisation Web Platform:

- A web-based GIS-like application providing:

  - Maps and plots for rainfall intensity, sea level, and hazards (pluvial and coastal).

  - Options to add flood mitigation measures such as barriers and simulate multiple mitigation scenarios.

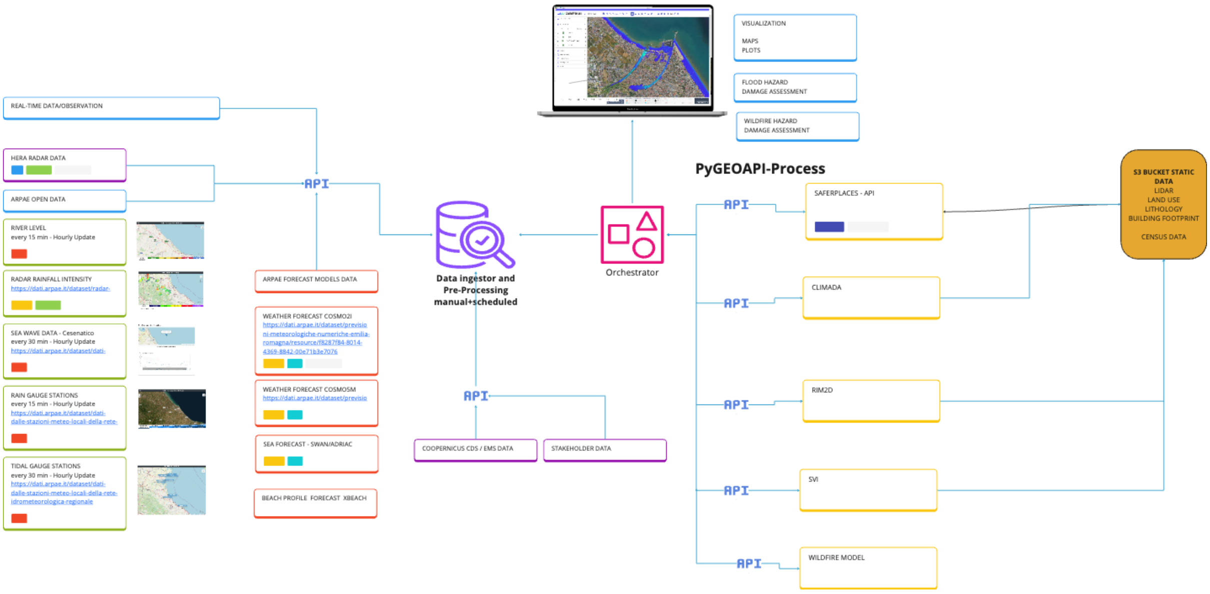

Figure 12: Flow chart of data, processes and results for the RWL Emilia Romagna Region.

The architecture addresses key user stories, such as those of civil protection officers and municipal authorities, by providing the necessary tools for real-time flood management, hazard mapping, and climate change adaptation in a user-friendly, interoperable platform.

Funded by the European Union

3.3 RWL 3 - Danube Region, Vienna, Austria, and Zala Region, Hungary

3.3.1 Vienna

Personas

Title	Climate Reinsurance Representative
Name	
Organisation	Hannover Reinsurance Company
Responsibility (in general)	To analyse data provided by platforms such as HORA to gain information on early-warning and potential hazard events for reinsurance claims

Title	Climate Insurance Representative
Name	
Organisation	UNIQA Insurance Group, Vienna
Responsibility (in general)	To look at climate forecasting models and scientific papers published to inform UNIQA on potential insurance claims related to pluvial and fluvial flooding in Vienna (with a focus on pluvial flooding, as the weir system in Vienna currently protects the city from overflows from the

	Danube)

Title	**Climate Scientist**
Name	
Organisation	Geosphere Austria
Responsibility (in general)	Providing up-to-date information on climate models and adapting modules within the Geosphere Austria platform

User Stories

Title	**HORA-like Tool for “Nearcasting” and Early Warning in the Danube River Basin**
Who?	Climate Reinsurer at Hannover Reinsurance Company
What?	As a climate reinsurer at Hannover Reinsurance Company, I analyse data provided by platforms such as HORA to inform the company about potential reinsurance claims. However, to gain this information, I oftentimes have to search through multiple different platforms and databases to get climate information about the entire river basin, especially with regards to extreme weather (i.e. convective storms and flooding events). As such, I need a centralised platform similar to HORA, that spans across multiple EU countries
Why?	As a reinsurance company, our jurisdiction spans across multiple EU countries, and thus we need information that spans across the Danube River Basin
How?	I would like to see a platform similar to HORA which includes weather forecasts and building asset damages at a city-level; having a dashboard/legend which also includes model comparison would also be quite interesting, as we do not communicate adequately with the science industry

	–which key target areas within the Danube River Basin

Title	****Climate Forecasts for Pluvial Flooding in Vienna**
Who?	Climate Insurer at UNIQA Insurance Group, Vienna
What?	As a climate insurer, I look at scientific papers and forecasting models to look at potential hazard events, particularly with regard to pluvial flooding. I need a way to centralise this information, because sifting through up-to-date scientific papers and forecasting models can be time-consuming and difficult to get a clear picture from
Why?	Centralised pluvial flood information would be very helpful for the company and which parts of Vienna could potentially be impacted by pluvial flooding events
How?	I would like a platform which gets information from sound climate models and can model pluvial flooding in the city of Vienna. I would like to see this information at a daily time-step, with forecasts in the next 72 hours, and future climate predictions for a 100 year flood statistic. I would like to be able to see building asset damages for all these forecasts. Ideally, these results would be displayed in an interactive WebGIS, and be able to print out a pdf of the flood maps for internal documentation

Title	****Centralised Platform for Information Sharing between the Insurance and Science Sectors in Austria**
Who?	Climate Scientist at Geosphere Austria
What?	As a climate scientist at Geosphere Austria, I look at climate models and adapt the platform so that our information is robust and up-to-date. However, a lot of the information that we release on the platform has little to no connection with the insurance industry, which makes it difficult to inform citizens (i.e., there is an information mismatch that makes communicating climate risk difficult). Thus, I need a way to communicate climate information with insurance companies in Austria

Why?	This would help to bridge the gap between science and the insurance industry for everyday citizens
How?	I would like to have a way to compare our climate models with the information the insurance industry uses, ideally in a centralized platform containing their climate model outputs and some of the ones that we host ourselves. I would also like a way to communicate with various actors in said platform

Title	**Drought Projections within the Danube River Basin**
Who?	(take out insurance companies by name)
What?	
Why?	
How?	

Data

[to be completed upon prioritisation and selection of user stories]

Components used from the Data Fabric

[to be completed upon prioritisation and selection of user stories]

3.3.2 Zala region

Personas

Title	**Member of the Water at Lake Balaton** **and/ or Mayor (of Keszthely)?**
Name	
Organisation	
Responsibility (in general)	Tourism safety and management at the Lake. Works on improving safety measures along the lake (lifeguards, response, medical, safety infrastructure) Link to insurance - drought and flooding at larger scale?

Title	**Member of the Zala Special Rescue Team** **/ other voluntary organisations in Zala + relationship with municipalities**
Name	
Organisation	
Responsibility (in general)	Responds to hazard events in the Zala Region, namely at a residential level Voluntary organisations are supporting local municipalities with disaster preparedness and response locally (e.g. sandbags, tree

	removal) and awareness raising (e.g. engagement in schools). This support is likely to increase and need expansion of more resources in the future given increasing frequency and intensity of multiple climate and weather-related hazards. Opportunity to learn from other European countries through the Real World Lab knowledge exchange, e.g. Emilia-Romagna and Denmark. Opportunity to explore future scenarios (explore impacts/ measures) in workshops (starting with flooding, evolve to multi-hazard) and discuss how as voluntary organisations they need to adapt and explore how the collaboration with municipalities could evolve to support DRR and CCA. Workshops focus on engagement of voluntary organisations (how can they become more resilient) and can be a starting point for deeper science-informed dialogue between municipalities and voluntary organisations on DRR/CCA. The risk communication products produced (e.g. risk maps, procedural workflows) can be used for this dialogue.

User Stories

[to be completed upon prioritisation and selection of user stories]

Title	**Volunteer organisations Future Climate Readiness / Resilience**
Who?	Voluntary member of the subunit of the Zala Special Rescue Team in the City of Keszthely, in the West Balaton Region.
What?	Support medium-term planning (5-10 years?) within and across all voluntary organisations for future extreme climate and weather-related events (fluvial and pluvial flooding, drought, wildfires) in the West Balaton region. For example, to understand what socio-economic impacts future flood events (at the city and west Balaton level) under different climate scenarios (e.g. for 2030 and 2050) will have on the city and west Balaton region and how this needs to then inform how voluntary organisations work.

Why?	Encourage more bottom-up community-led Civil Protection connecting Disaster Risk Reduction and Climate Change Adaptation through stronger collaboration between voluntary organisations in the West Balaton region to strengthen climate and disaster resilience. Understand how voluntary organisations can best prepare for and adapt to the impact of climate change on their emergency preparedness and response planning and actions, as well as their climate/disaster related outreach and awareness activities with the public/ citizens. For example, understand if synergies can be captured through combining or sharing resources in the most impacted areas. Facilitate stronger collaboration with municipalities (to advocate for additional resources/ needs) through enhanced access to risk communication products (e.g. risk maps).
How?	Through collaborative planning workshops with members from different voluntary organisations co-exploring different future scenarios and exploring adaptation/ preparedness measures and actions. Flood model at city scale (SaferPlaces, Rim2D) and West Balaton scale (Danube model - tbc?) Precipitation projections West Balaton scale - (PIK climate models/ Danube model?) Wildfires - TBC Climada - impact estimations / multi-criteria analysis for different future scenarios

Title	**West Balaton Regions Future Climate/ Disaster Resilience**
Who?	Mayor of the City of Keszthely (one of up to 100 mayors in the West Balaton Region)
What?	Ensure their city and region is prepared for and can adapt to future extreme weather and climate extreme events and socio-economic impacts (flooding, drought, wildfires) with a focus on what they can do in their election cycle (5

	years).
Why?	To strengthen climate resilience of citizens and regional development sectors (economic growth and tourism) Identify pathways to action - measures, resource distribution, governance
How?	Risk maps/ simulations for future flooding (pluvial and fluvial) and future precipitation/ temperatures (drought, wildfires) in City of Keszthely and West Balaton region (where possible) Real World Lab Learning exchange with examples (risk products/outputs) from other cities/ regions (e.g. Denmark and Emilia-Romagna) Collaborative workshop with mayors in the region and also bilateral exchange directly with specific mayors to showcase specific results. Model chain/ outputs would be the same as the volunteer organisations
Title	**Long-term climate/ drought adaptation planning** **(lower priority/ nice to have)**
Who?	Land managers responsible for long-term planning of activities (10, 50, 100 years) e.g. forest owner, orchid owner, farmers in the West Balaton region.
What?	Understand how precipitation will change in the next 50 to 100 years and if there will be more frequent/ severe droughts
Why?	Identify how land managers should adapt their practices e.g. planting of certain types of trees (need to aim to import from abroad), investing in an irrigation system

How?	Workshop with land use managers Precipitation modelling with climate scenarios - PIK? Clarify what Tobias can produce from his models. Climada - something on uncertainty?

Title	**Seasonal Precipitation Forecasts at Lake Balaton**
Who?	Water authorities?
What?	As a member of (xx organisation), I would like more information on wet and dry seasons at the lake, as this is not currently available
Why?	This would help with planning efforts, especially in the tourism industry. If we know the seasonal forecasts, we would know what kind of safety measures we need to prepare for (i.e., in wet seasons we need sandbags, in dry seasons we need to order more shade and wildfire safety precautions as this has been an issue in the past)
How?	I would like to see seasonal forecasts for precipitation at the lake, to know mainly when we are facing dry seasons. I would like these forecasts to be 1 month in advance, and be able to see where particularly hotspots along the lake might be located.

Title	**Drought modelling**
Who?	
What?	

Why?	
How?	Would like to see in 5-10 years, maybe 50, droughts in the region –maybe more generalised agricultural losses? (i.e. what kind of agricultural output is Tobias producing? What can it show?)

Title	**Residential Flood-Risk Map in the Zala Region**
Who?	Member of the Zala Special Rescue Team
What?	As a member of the Zala Special Rescue Team, I would like to know which neighbourhoods to respond/give priority to in times of flooding events.
Why?	This would help my team and I in our organisation, and where to coordinate our efforts to maximise safety and minimise risks to citizens under our jurisdiction
How?	I would like to see a flood map in real or near real-time generated in high predicted precipitation events. I would also like warnings to be issued within the Data Fabric platform for this use-case. If possible, I would like to monitor precipitation (even in non-flooding times) at a real or near real-time daily time-step.

Data

[to be completed upon prioritisation and selection of user stories]

Components used from the Data Fabric

[to be completed upon prioritisation and selection of user stories]

3.4 RWL4-Rhine-Erft Region, Germany

Personas

Title	**Erftverband's employ**
Name	
Organization	
Responsibility (in general)	

User Stories

Title	**20% Increase of Green Roof Flood Simulation at the Past, Present, and Future**
Who?	
What?	
Why?	
How?	

Title	**1L Tanks Flood Simulation at the Past, Present, and Future**

Who?	
What?	
Why?	
How?	

Title	**1L Tanks and Green Roofs Flood Simulation(s) with Past Precipitation/Flooding Event Levels**
Who?	
What?	
Why?	
How?	

Title	**HOWIS User-Friendly Cloud-based Data Storage and Visualization**
Who?	Erftverband
What?	Improvement of the data fed into the HOWIS platform to make it easier to understand and more user-friendly

Funded by the European Union

Why?	The current data present in the HOWIS platform is too complex for Erftverband employees to interpret and understand in a meaningful way; additionally, data in times of disaster events is from a plethora of sources, making it difficult to get a quick overview of the event/hazardous rainfall and weather forecasts (from DWD, the Ministry of Environment, Nature Conservation and Transport of the State of North Rhine-Westphalia, the State Agency for Nature, Environment and Consumer Protection of North Rhine-Westphalia, as well as water boards). Thus, an easily accessible and centralised cloud-based storage to analyse and visualise data would be beneficial to support flood-risk management.
How?	

Title	**Digital Twin**
Who?	
What?	
Why?	
How?	

Title	**Agricultural Modelling (discussed in modellers' meeting)**
Who?	
What?	
Why?	

How?	

Data

[to be completed upon prioritisation and selection of user stories]

- HOWIS data ingestion into OGC Connected Systems API

Components used from the Data Fabric

[to be completed upon prioritisation and selection of user stories]

Funded by the European Union

4 Conclusion

This deliverable details the low-level architecture of the DIRECTED Data Fabric. The architecture is designed in a modular and federated way. Technical components have been described and necessary interfaces have been identified. The driving requirements stem from the discussions with the RWL hosts and their stakeholders as summarised in personas and user stories in Section 3. The components constituting the core Data Fabric are selected from proven Open Source Solutions as well for the authorization and access control as well as for the Spatial Data Infrastructure managing the data in the Data Fabric. Models will be interoperably integrated where possible. For some closed source solutions, only the pre-defined model results and outputs can be integrated. Complexity of data discovery and data access and orchestration of processes is hidden from the user and encapsulated in web-based front-end components. Additionally, the Data Fabric offers APIs to retrieve data and results through existing Web and desktop GIS applications.

Based on the defined architecture, the implementation of the modular components of the Data Fabric can now be carried out. In order to ensure that the developments will continue to address the needs and requirements of the users in the RWL, iterative reviews of the solution with RWL hosts and stakeholders will be established. This co-development phase is i.e., important for the design and development of the customised information products, which for some RWL will be a dedicated web application and for others a layer that is imported into existing solutions. Hence, the Data Fabric will showcase how improved interoperability contributes to DRM and CCA.

4.1 Constraints

The Data Fabric is designed to support open standards and to be as flexible as possible. However, only a selection of different data sets and models are exemplarily integrated in the prototype of the Data Fabric. This does not guarantee full flexibility to integrate virtually every model or data set. Description of the processes (i.e., wrapped models) and metadata of processes, models and data are provided guiding the user in integrating new data and setting up new model chain runs.

The Data Fabric is a tool provided to assess risks and impacts of natural hazards based on the integrated models. The responsible use of the Data Fabric is supported with guides and questionnaires (see also Deliverable D5.3 and D5.4). However, this cannot be enforced by the Data Fabric as such, but the Data Fabric will document model runs in order to provide means to scrutinise and audit results generated by the Data Fabric.

The Data Fabric developed within DIRECTED is a highly modular system of mainly open-source components. Along with the software, there will be scripts to deploy the Data Fabric in cloud environments. The operation of the Data Fabric beyond DIRECTED can in general also be achieved in subsets.

4.2 Licences

The main components of the Data Fabric rely on open-source solutions and as such will the Data Fabric also adhere to the individual licences of used libraries and solutions. Some models are however not published under an open-source licence and only model results can be used in the Data Fabric - not the model on request.

Besides the software, also data has a licence for usage. Information will be requested during the integration of data sets from the data steward entering new data. For model runs based on input data, a suitable licence will be suggested and needs to be confirmed by the creator of the model chain generating those results.

Licensing will not change after the DIRECTED project, but access granted within DIRECTED to dedicated software solutions of partners might be retained and individual agreements between RWL and software providers need to be made.

Funded by the European Union

Annex 1

Partners

Funded by the European Union

This project is an Innovation Action under the Civil Security for Society, Disaster-Resilient Societies programme of the Horizon Europe funded by the European Union.
Associate partners SEI Oxford and Oasis Hub are funded by Innovate UK and ETH Zurich is funded by The State Secretariat for Education, Research and Innovation (SERI), Switzerland