Open
Geospatial
Consortium

OGC DISASTER PILOT 2021 ENGINEERING REPORT

ENGINEERING REPORT

PUBLISHED

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I   EXECUTIVE SUMMARY

A disaster, whether anticipated or not, can be an overwhelming event. Each disaster has its own unique details and progression, which can generate and cascade into additional crisis situations. The preparation and coordination of scalable responses can mitigate these challenges. Disaster Pilot 2021 brings the technological pieces together in order to reduce information preparation time and accelerate the ability to transform data from observation into decision. This will require bridging the divides between providers, responders and other stakeholders, forming a connected ecosystem of data and technologies, and developing the capacity to produce DRI (Decision Ready Information) products that answer decision makers' questions almost as fast as they can be posed.

In order to successfully implement and test critical technical and organizational aspects of the envisioned ecosystem, the Pilot focused on scenarios addressing hazard early warning, monitoring, vulnerability assessment, disaster detection, impact awareness, and disaster response related to:

- Landslide / flooding hazards and pandemic impacts within the Rímac and Piura river basins in Peru;

- Flooding hazards and pandemic impacts within the Red River basin in Manitoba, Canada; and

- Integration of Health and EO data and services for pandemic response the greater New Orleans area in the United States.

The Pilot was developed through stakeholder consultation in multiple workshops and meetings. It included DRI recipes that integrated earth observations with social, economic, health, infrastructure, environmental, and other information to address key indicators of disaster risk, vulnerability, and impact with the goal of directly influencing the scope and nature of disaster and pandemic response activities.

Further Pilot activities investigated getting this right information into the hands of the right people on the ground. For response practitioners, this involved generation of secure, portable Geopackage archives for distribution in connectivity-challenged circumstances. For the public, the Pilot explored Web publication of disaster information Web pages including linked structured data in order to connect well-known local geography with up-to-date conditions, observations, and predictions in common Web searches.

II   KEYWORDS

The following are keywords to be used by search engines and document catalogues.

ogcdoc, OGC document, JSON

III PREFACE

It is possible that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed upon by any implementation of the standard set forth in this document and to provide supporting documentation.

IV   SECURITY CONSIDERATIONS

The overall aspect of security, authenticity, integrity, and confidentiality are tightly coupled, not only in emergency and disaster management situations. Ensuring trust in data being used for disaster management is a major concern. In terms of establishing trust into the data via security capabilities provided with online communication (e.g. HTTPS) does not apply, as the security mechanisms must be attached to the data itself to support the essential offline use. In OGC Testbeds since #15, this is defined as Data Centric Security.

The efficient and effective management of disaster impacts relies on many factors, including the interoperable access to relevant data provided by different stakeholders in existing Spatial Data Infrastructures (SDIs). The relevant data that otherwise might not be accessible, must be made available seamlessly to all identified actors of disaster management ensuring security and trust. The OGC GeoDRM Abstract Specification defines this ability as overriding access policies by **pushing the red button**.

But, enabling the "red button override" must not necessarily result into the loss of authenticity, integrity, and confidentiality of the data. As a stakeholder, it is important to be able to limit the further use of the data released for the purpose of disaster management. Constraints certainly involve the role of actors, also time and spatiotemporal constraints: who, when, and where shall access to the data for a designated disaster area be allowed? Extending the concepts of GeoXACML and Data Centric Security, explored in OGC Testbeds #15, #16 and the ongoing #17, helps to define solutions for maintaining security when packaging encrypted data for controlled use with verifiable trust as it is key for the actors.

During normal operations, system security controls guarantee GDPR compliance regarding personal data. But once the "red button override" is active, GDPR relevant consent buttons might countermeasure the effectiveness of responders. One typical solution to make data quickly available for disaster response management is to drop access control and put in place extra logging and provenance recording. But for online access, this approach introduces the threat of data leaking and can therefore not be used. Also, disabling access control would also have implications to data privacy and ethics.

Linking SDIs with Health infrastructures is a new emerging concept that is extremely relevant since the COVID-19 impact demonstrated its importance. Ethical concerns and data privacy are extremely relevant, but how can this be achieved under disaster management circumstances? From a technical perspective, it would potentially be too difficult for each system that provides disaster management data to be updated with an interoperable mechanism that supports the "red button overwrite". Instead of trying to change existing APIs and their security controls, the concept was developed during Disaster Pilot '21 where stakeholders produce a secure, trustworthy container of data that ensures integrity and confidentiality but also supports trust. The proposed solution is based on Secure GeoPackage that can be produced by stakeholders and then shared with relevant actors.

V SUBMITTING ORGANIZATIONS

The following organizations submitted this Document to the Open Geospatial Consortium (OGC):

- Secure Dimensions GmbH

VI SUBMITTERS

All questions regarding this document should be directed to the editor or the contributors:

NAME	ORGANIZATION	ROLE
Andreas Matheus	Secure Dimensions	Editor
Tony San José	SATELLOGIC	Contributor
Ignacio Correas	Skymantics	Contributor
Dave Jones	Stormcenter	Contributor
Zhe Fang	Wuhan University	Contributor
Simone Giannecchini	Geosolutions	Contributor
Paul Churchyard	HSR	Contributor
Guy Schumann	RSS-Hydro	Contributor
Dean Hintz et.al.	Safe Software	Contributor
Omar Barrilero	SatCen	Contributor
Josh Lieberman	OGC	Contributor
Pedro Gonçalves	Terradue	Contributor
Fabrice Brito	Terradue	Contributor

1

SCOPE

1   SCOPE

This report is a summary of activities undertaken in the execution of Disaster Pilot 2021. These included integration and transformation of EO provider data using hybrid applications-to-the-data EO cloud exploitation platforms to seamlessly bring analysis-ready imagery, in situ, social, economic, environmental, health, and other data streams into scalable cloud environments where advanced processing, modeling, and algorithms can be directly and flexibly applied to them. They also included assessing and validating ARD standards, then integrating both space-based and local data sources on demand, providing targeted decision ready information products to local analysts and field responders through modern convenience API's, optimized hybrid-cloud services, and mobile-ready online-offline Geopackage tools. Additional activities included Web search optimization for disasters through Web publication of linked "structured data" connecting well-known local geography with up-to-date conditions, observations, and predictions. Pilot activities focused on pandemics, flooding, and landslide disasters in the Piura and Rímac river basins of Peru, the Red River basin of Manitoba, and the greater New Orleans area.

2

TERMS, DEFINITIONS AND ABBREVIATED TERMS

———

2 TERMS, DEFINITIONS AND ABBREVIATED TERMS

This document uses the terms defined in OGC Policy Directive 49, which is based on the ISO/IEC Directives, Part 2, Rules for the structure and drafting of International Standards. In particular, the word "shall" (not "must") is the verb form used to indicate a requirement to be strictly followed to conform to this document and OGC documents do not use the equivalent phrases in the ISO/IEC Directives, Part 2.

This document also uses terms defined in the OGC Standard for Modular specifications (OGC 08-131r3), also known as the 'ModSpec'. The definitions of terms such as standard, specification, requirement, and conformance test are provided in the ModSpec.

For the purposes of this document, the following additional terms and definitions apply.

2.1. Terms and definitions

2.1.1. application

A self-contained set of operations to be performed, typically to achieve a desired data manipulation, written in a specific language (e.g. Python, R, Java, C++, C#, IDL).

2.1.2. application package

A platform independent and self-contained representation of an Application, providing executables, metadata and dependencies such that it can be deployed to and executed within an Exploitation Platform.

2.1.3. compute platform

The Platform providing the computational resources for the execution of the Application.

2.1.4. container

A container is a standard unit of software that packages up code and all its dependencies so that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

2.1.5. coordinate reference system

coordinate system that is related to the real world by a datum term name (source: ISO 19111)

2.1.6. exploitation platform

An on-line system made of products, services and tools for exploitation of data.

2.1.7. portrayal

presentation of information to humans (source: ISO 19117)

2.1.8. LiDAR

Light Detection and Ranging — a common method for acquiring point clouds through aerial, terrestrial, and mobile acquisition methods.

2.1.9. DCS

Data Centric Security

2.1.10. **KMS**

Key Management System

2.1.11. **Decision Ready Information Component | D111-3 | Secure GeoPackage Service (aka GeoPackage Encryption Service)**

Web Service extension to WMS 1.1 and WFS 2.0 to produce GeoPackages with encrypted content upon request

2.2. Abbreviated terms

API	Application Programming Interface
COG	Cloud Optimized GeoTIFF
COM	Component Object Model
CORBA	Common Object Request Broker Architecture
COTS	Commercial Off The Shelf
CWL	Common Workflow Language
DCE	Distributed Computing Environment
DCOM	Distributed Component Object Model
DCS	Data Centric Security
DEK	Data Encryption Key
EGS	Encrypted GeoPackage Service
EO	Earth Observation
GDAL	Geospatial Data Abstraction Library
IDL	Interface Definition Language
JSON	JavaScript Object Notation

KEK   Key Encryption Key

KMS   Key Management System

REST   Representational State Transfer

STAC   SpatioTemporal Asset Catalog

URL   Uniform Resource Locator

3

ABSTRACT

3 ABSTRACT

This OGC Disaster Pilot '21 (DP21) Engineering Report summarizes work done in the Pilot to increase disaster awareness among a range of disaster management stakeholders. Pilot participants implemented components of a data flow ecosystem to leverage analysis-ready earth observations and other datasets (ARD) and produce decision ready indicators (DRI) according to collaboratively developed workflow recipes. DP21 focused on the hazards of flooding, landslides, and pandemic, as well as the interactions and complications between them, in three regions including the Piura and Rimac river basins in Peru; the Red River Basin in Manitoba, Canada; and the greater New Orleans area in Louisiana, United States. The Pilot also prototyped providing information to field practitioners in secure geopackage formats, as well as leveraging linked data and structured web page information to optimize public web searches for disaster information.

4

INTRODUCTION

---

4 INTRODUCTION

When a disaster strikes it can generate additional crisis situations. These can be overwhelming, but preparation and coordination of scalable response capacity can meet the challenges. Disaster relief forces from supporting jurisdictions can quickly integrate and analyze vast streams of realtime data from multiple sources to monitor the evolving situation and plan appropriate responses. Hybrid scalable cloud-based systems bring advanced AI processing, machine learning algorithms, and predictive models right to where earth observation and other data is already uplinked, prepared, and curated, generating analysis-ready data (ARD) with the characteristics, scale, and speed that the complex situation demands. Convenience API's, as well as download-and-go geopackage containers and viewers, bring decision-ready information (DRI) directly to field workers' mobile devices even in resource-constrained, low-connectivity areas. Meanwhile, web publication of "structured data" that link well-known content with up-to-the-minute situation observations enables search engines to push disaster-relevant information up in search results and help the public stay on top of fast-moving events. All of this is possible through the preparation of technologies, geospatial standards, and data sharing arrangements that bring the right information at the right time to the right people in the right place.

Following the success of Disaster Pilot 2019, the OGC Disaster Pilot 2021 brought all the above mentioned technologies together. Though elements had been explored and tested in various scenarios, it is OGC's unique position in the center of the geospatial marketplace that allowed a single pilot to explore and test key state-of-the art technologies in an unprecedented combination. The Pilot was able to serve as an accelerated market research activity, while at the same time demonstrating the integration potential native to OGC Innovation Program activities. The emphasis was on testing end-to-end information flow related to the most urgent phases of disaster management, with an emphasis on first responders and other end users. The Pilot focused on four essential elements.

1. Scalable exploitation of earth observation data on hybrid cloud platforms using application-to-the-data technologies.

2. Development and implementation of recipes integrating and transforming ARD into DRI to provide specific guidance for decision and actions during disaster preparation, response, and recovery.

3. Generation of secure GeoPackage offline containers to facilitate taking all relevant information into the field.

4. Capability for structured data to enable web search engines to push disaster-relevant information up in search results during an ongoing disaster.

This work supported a vision of disaster response in which disaster relief forces are able to rely on information drawn from many EO and other data sources that can be integrated and analyzed quickly using the scalable capacity of hybrid, interoperable cloud platforms, specific adaptable recipes for generating decision ready indicators, and the power of models to distill relevant information into specific predictions, for example, flood risk.

The Pilot focused on scenarios addressing hazard early warning, monitoring, vulnerability assessment, disaster detection, impact awareness, and disaster response related to:

- Landslide / flooding hazards and pandemic impacts within the Rímac and Piura river basins in Peru;

- Flooding hazards and pandemic impacts within the Red River basin in Manitoba, Canada; and

- Integration of Health and EO data and services for pandemic response in the greater New Orleans area in Louisiana, United States.

4.1. Previous Work

The OGC Innovation Program has included many initiatives over more than two decades addressing spatial data sharing for disaster management, emergency response, situation awareness, and other critical action scenarios. More recent relevant activities include the following.

- Incident Management Information Sharing Pilot (IMIS) (2016) applied OGC principles and practices for collaborative development to existing standards and technology to prototype an IoT approach to sensor use for incident management and situation awareness.

- Disasters Interoperability Concept Development Study (2018) focused on understanding how best to support the development or integration of SDI(s) for use in disasters.

- Disasters Pilot 2019 (2019) focused on the demonstration of the usefulness of standards and SDI architecture within the Disaster community.

- EO Cloud Platform CDS (2021) evaluated EO cloud platform architectures and alignment with open standards, as well as documentation of their readiness to support the Disaster Pilot 2021 disaster response exercise scenarios.

- Health Spatial Data Infrastructure CDS (2021) drafted a common, standardized health geospatial data model and schema to establish a blueprint for better aligning and preparing the community for early warning, response to, and recovery from future health emergencies.

4.2. Analysis Ready Data

By some estimates, 80% of response time is spent procuring and preparing data, leaving only 20% to be spent learning from and acting on it. The more that this data preparation can be carried out in advance of data needs, the more that urgent time can be spent benefitting from the data itself. For this reason, Analysis Ready Data (ARD) is vital to support immediate

responder decision-making. Many sources of data, however, particularly EO data, can be hard to find, complicated to share, difficult to access, and slow (or impossible) to process into common forms that are suitable for analysis and integration.

The most recognized definition of ARD is from the Committee on Earth Observation Satellites (CEOS): *satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets...The definition of CEOS-ARD reflects the attributes of fundamental measurement products for the majority of global remote sensing users with land imaging applications, and are the minimum level required to support time series analysis and data interoperability.*

CEOS have developed and are developing a number of ARDs Product Family Specifications for common EO types, such as Surface Reflection for optical imagery and Normalized Radar Backscatter for SAR imagery. There are a number of challenges to this specification process, besides the large and growing number of varieties of earth observation data including the following.

- The type of analysis to be performed and/or the application for which the data is to be used have an impact on what is needed for data to be analysis ready.

- The concept of readiness for whatever is the next step of analysis applies not only to "raw" earth observations, but also to any intermediate steps between collection and interpretation, including outputs from models such as predictive or classification algorithms, as well as other types of spatial data to be integrated, such as framework, reporting, or statistical data.

In Disaster Pilot 2021, a broader but less specific definition of ARD is developed, namely readiness or close to readiness of any dataset for broad categories of the next step in transforming and integrating that dataset for applications to improving disaster awareness. The concept of a geospatial data cube is useful here, in that datasets having a common definition of orthogonal space, time, and phenomenon axes are more likely able to be compared or combined for better coverage of a data space. An example would be combining EO data from multiple instruments for a longer time series or better cloud-free continuity. Another example could be combining observations or predictions made at different resolutions to provide a "drill-down" perspective. The point here is that the initial development of ARD concepts has proven very useful, but there is much more that could be done to reduce data "wrangling" for a broad range of datasets and analytical applications.

4.3. Clouds

Cloud computing is the capability to orchestrate remote computational workflows performed by essentially unlimited numbers of virtual processors, data stores, load balancers, and other processing components. Made possible by vast regional data centers containing hundreds of thousands of physical components and voluminous Internet bandwidth, cloud computing can reduce weeks or months of computational hardware acquisition and installation to a few minutes of laptop dashboard time. One of the few limiting factors in cloud computing

capabilities is external: the cost, time, and difficulty of moving data into or out of a particular cloud computing environment or platform.

The cloud computing paradigm has a direct bearing on the use of EO data, satellite based or otherwise, for disaster awareness. The volume of Earth Observation (EO) has drastically increased in recent years, making it more likely that a particular disaster response can benefit from collected observations. An increasing number of satellites and improved capabilities (better spatial and temporal resolution) of new imaging sensors, as well as improved weather and other predictive models fed by them, have led to an exponential growth in volume, velocity, and variety of the daily EO data stream. EO data providers and users alike are faced with challenges in managing, processing, and handling this growth. Such volumes of EO data can no longer be downloaded and processed on local machines at either the necessary scales or speeds, nor for the variety of applications needed.

The cloud computing paradigm is able to meet these new requirements. As long as the raw observations can be efficiently uploaded to a cloud data center, new cloud-based EO exploitation platform data analysis workflows can then focus on bringing the much smaller scale application codes to where that data is stored. Cloud-based computing services can then provide virtual storage and computing resources that are deployed as needed and disposed of just as soon as they are no longer needed, so that overall IT costs for working with EO data can be significantly reduced.

As described in more detail in the EO Clouds Concept Development Study Report, this cloud computing applications-to-the-data paradigm is highly promising, and also highly disruptive to traditional concepts of data possession. As many of the participant component descriptions can attest, access to cloud computing can make all the difference in responding to urgent disaster awareness needs, but only if challenges can be overcome in producing and accessing analysis ready data, as well as delivering useful products from cloud workflows to people in disaster-impacted regions. Other challenges arise in adapting application codes to the specific cloud platform environment where input data are stored, or where multiple input datasets are stored on different platforms or regions.

The cloud is important but a great deal of technical, cultural, political, and educational work needs to be done to make it sufficient for improved disaster management based on earth observations.

4.4. Decision Ready Indicator Recipes

It has long been recognized, and is specifically called out in the Disaster Pilot 2021 CfP, that EO data need to be adapted — analyzed, integrated, and interpreted — in order to be of use in guiding the actions of disaster management stakeholders. As expressed in the CfP: "to bridge this divide, bringing together the Disaster SDI puzzle pieces that connect the right players from the data providers all the way to the first responders and the decision makers and everyone in between, forming a pattern that can adapt to any disaster, any region, any combination of data sources and tools." This was seen as a way to "accelerate our ability to transform data from observation into decision...bridging the divides between providers, responders and other stakeholders, forming a connected ecosystem of data and technologies, and developing the

capacity to produce Decision Ready Indicator (DRI) products that answer decision makers' questions almost as fast as they can be posed."

Through discussion and consultation between the Pilot participants and engaged stakeholders, a clearer concept has been developed for both the nature of DRI and the steps needed to generate DRI products from ARD sources.

- There is a fundamental transformation between ARD that represents observations (physical, social, and computational) of geographic features in the world, and DRI that represents guidance on actions that stakeholders may choose and/or pursue.

- The development of useful "recipes" for DRI products necessitates collaboration between multiple stakeholders as providers of ARD are unlikely to know well the particular decisions that need to be made, and decision makers are unlikely to know well how particular dataset can be used to guide them.

In the figure below are shown six representative user personas or roles for stakeholders in any ARD→DRI ecosystem. Responders and EO data providers are somewhat at each end of a DRI workflow. There are others, either in the middle such as analysts, or those ultimately affected by the result, such as the public. Each persona represents a unique set of responsibilities, knowledge, and interaction with a disaster information pipeline constituting an important role in devising DRI that are both feasible and effective.
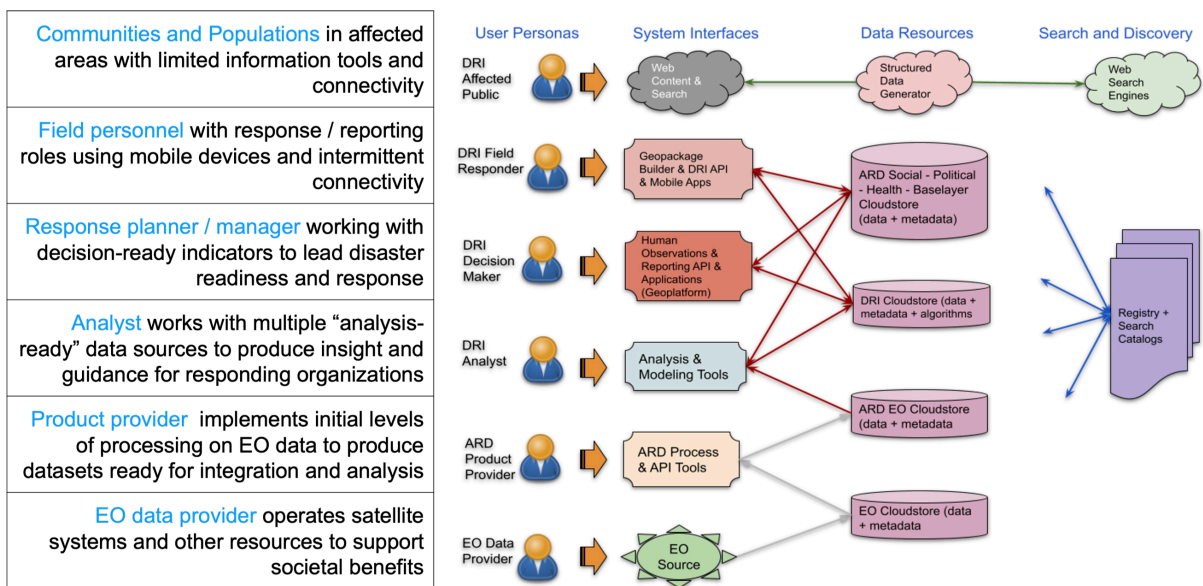
Figure 1 — Representative stakeholders in DRI recipe collaboration.

Assigning these roles to the the process of developing and implementing DRI recipes results in the idealized DRI collaboration sequence in the figure below. The process proceeds stepwise from observations of the world to specific types of indicators (e.g. detectors, predictors) relevant to decisions on specific hazards, actions, or impacts, with specific advice (e.g. be aware, get prepared, beware and take action). Ultimately, remotely sensed EO data are recognized as taking the place of infeasible or impossible local observations which might have informed responders of local needs.

The concept of in-person collaboration in the form of emergency mapping and analysis centers was raised repeatedly during the course of the Pilot and in many ways the DRI collaboration pipeline below is a way of recognizing this collaboration in a virtual distributed fashion.
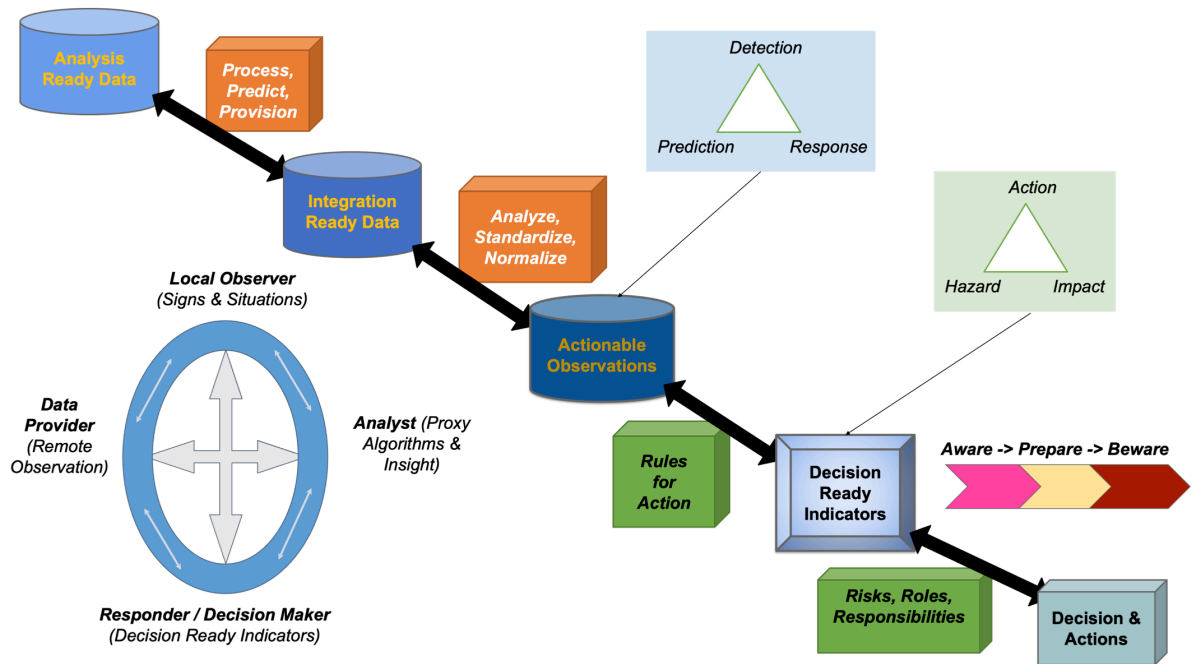
Figure 2 — Collaboration pipeline for a DRI recipe.

Ultimately, any distributed mapping and analysis center needs to develop the same trust in relevant DRI based on fit-for-purpose ARD that in-person exchanges are effective for. The figure below illustrates some of the components of that trust, many of which are addressed by components deployed and described below by Pilot participants.
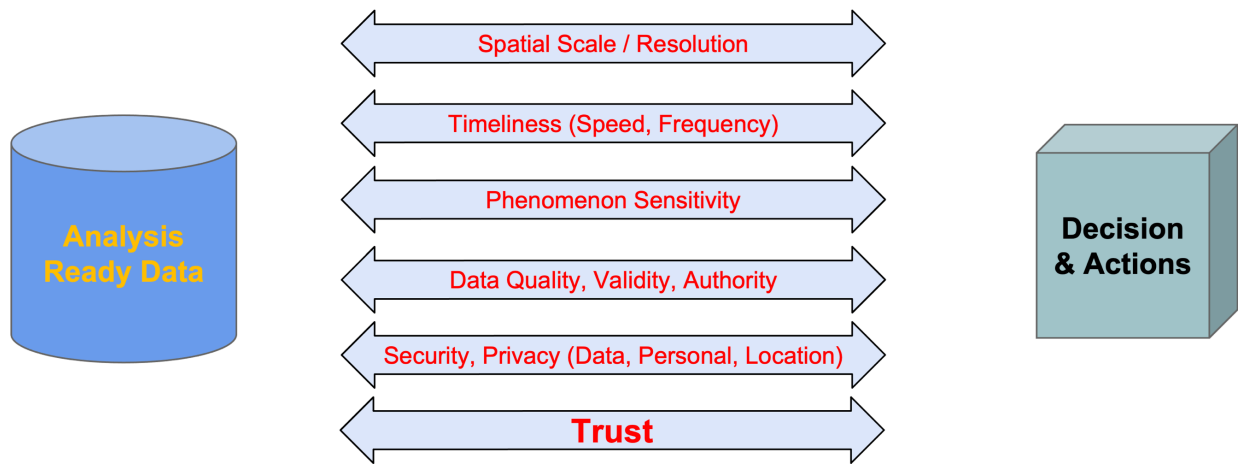
Figure 3 — Bases for trust in effective DRI recipes.

Disaster Pilot 2021 made progress in developing and testing the concept of DRI recipes, as well as implementing a number of recipes on a prototype basis for the 3 Pilot study areas, still recognizing that some balance will be needed going forward between the many possible permutations of DRI that could be defined and those that are the most important for good decisions in the widest range of circumstances.

4.5. GeoPackage Security

The use of Secure GeoPackage allows the establishment of new disaster management relevant data distribution opportunities that are not possible using existing SDI. Multiple pros can be identified when using Secure GeoPackage.

- The creation of Secure GeoPackage can be done using the existing SDI data and security controls.

- A Secure GeoPackage contains a well defined set of disaster management relevant data that can be trusted.

- The roll-out of the Secure GeoPackage(s) can take place via regular file sharing, e.g., stored on mobile devices, shared network drives, or via WebDAV.

- Trust in a Secure GeoPackage is enabled by ensuring:

  - confidentiality of the data using encryption;

  - integrity via digital signatures;

  - authenticity by evaluating the identity behind the digital signature; and

  - access control by managing access to data (decryption) keys via Key Management System.

- The interoperable data structure of a Secure GeoPackage ensures backwards compatibility with existing implementations. The Secure GeoPackage can be loaded but for rendering the encrypted data, additional client functionality is required.

- Generation of Secure GeoPackage can be implemented as a dedicated API that can be deployed in addition to existing stakeholders services.

4.5.1. Secure GeoPackage

In an ideal interoperable world, access to APIs and data could be switched from normal to emergency use as required by disaster management. However, switching a Zero Trust domain's security controls from normal to disaster management operation is almost impractical as it would introduce threats towards data leakage.

One alternative approach that was studied, implemented, and deployed in the Disaster Pilot '21 focused on the use of GeoPackage with encrypted content — aka Secure GeoPackage. The encryption ensures confidentiality of the data by introducing a Key Management System. An interoperable and flexible infrastructure was deployed that demonstrates the packaging of disaster management data honoring the existing (provider) specific security controls but still ensure emergency use access to the encrypted data of a Secure GeoPackage.

A Secure GeoPackage defines — as a GeoPackage extension — a particular data structure that slightly differs from the standard GeoPackage structure: The Secure GeoPackage contains a table to store the encryption key(s) metadata and the data tables include one additional column that is used as a foreign key to the encryption key's metadata. The Secure GeoPackage ensures confidentiality, as the actual data is encrypted, but, the Secure GeoPackage itself does not provide integrity and authenticity.

Integrity of a Secure GeoPackage can be established by introducing a secure site like a Content Distribution Network (CDN) that allows registration of the hash for the GeoPackage based on its filename. For checking integrity, an application can verify the integrity of the Secure GeoPackage by comparing the self-generated hash with the hash obtained from the CDN for the filename. In case the CDN also offers to return authenticated hashes (digitally signed hashes), the app can further validate the authenticity of the Secure GeoPackage.

A Secure GeoPackage is an ideal container to store any kind of information relevant for disaster management because it does not require online access to the data when in the field. The secure GeoPackage(s) can be stored on mobile devices. Ensured by the Key Management System, controlled access to decrypt the data can be controlled without changing existing Spatial Data Infrastructures.
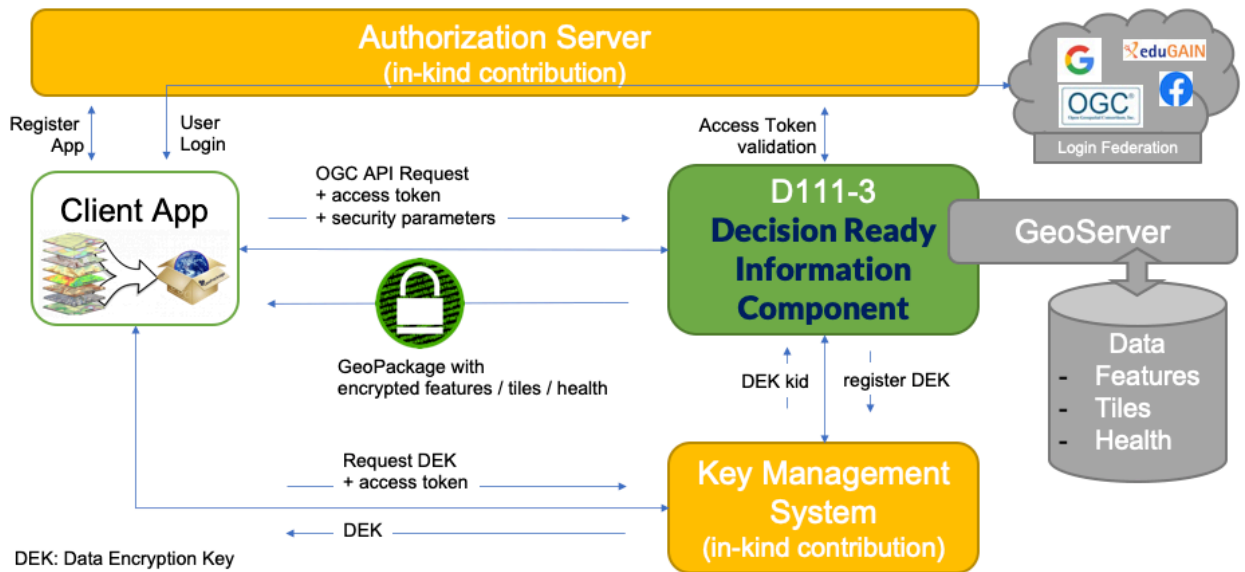
Figure 4 — Secure GeoPackage Component Overview and their messaging

For the Disaster Pilot '21, two GeoPackage extensions were defined for storage of encrypted (i) tiles and (ii) features. Both extensions are backwards-compatible with existing GeoPackage Reader implementations to ensure easy integration into broadly used GeoPackage clients like QGIS.
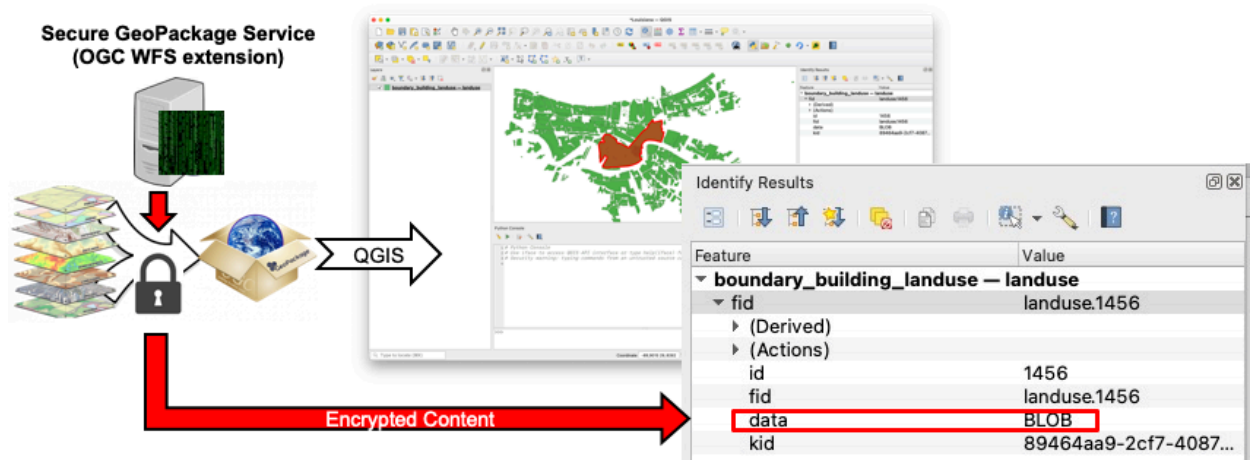
Figure 5 — Secure GeoPackage loaded in QGIS

The Secure GeoPackage can be loaded into QGIS as the structure is backwards-compatible, but, the user cannot see the encrypted content — it is displayed as BLOB.
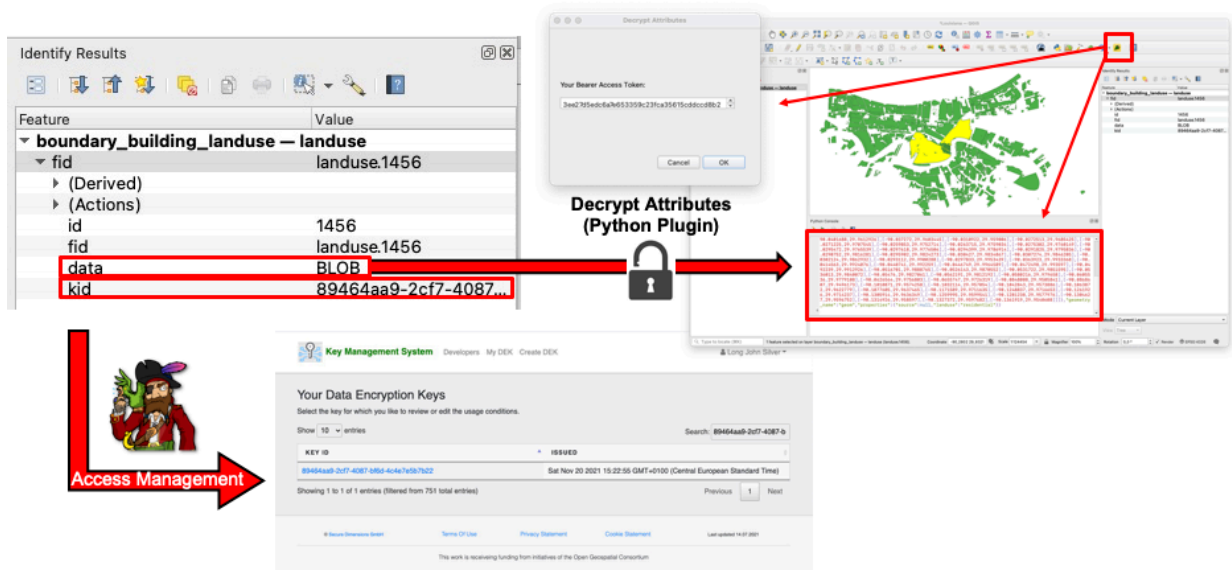
Figure 6 — Secure GeoPackage content decrypted in QGIS

The Secure GeoPackage encrypted data can be decrypted by loading the following Python plugins.

- Decrypt Secure GeoPackage Features

- Decrypt Secure GeoPackage Tiles

Figure 7 — Secure GeoPackage plugins for QGIS

The prototype implementations are available from Github: https://github.com/securedimensions/geoserver-geopackage-encryption-plugin

4.5.2. GeoPackage Encryption Extensions

The details for the GeoPackage extension can be found in the Annex.

4.6. Linked Data

Linked data is a common term for information representations and encodings in which much of the content is carried not by standalone data elements but by the relations or links between elements. This has affinities to semantic data representations which consist of logical statements defining a specific relationship (predicate) between one concept (subject) and another (object). The concept of linked data can also be seen in graphical forms of information representation

which may or may not carry the same axiomatic weight as a language supporting specific logics such as Ontology Web Language (OWL).

Linked data carries perhaps the most significance in relation to the connected nature of the Web, since the structure and organization of URL's provides a way not only to uniquely identify a data element but also to resolve or "dereference" the relationship between one element and another. For this reason, linked data representations are most closely connected with use of the Web as an information repository and Web-friendly formats such as JSON-LD.

In Disaster Pilot 2021, a specific role for linked data has been investigated, namely enhancing disaster information Web content via "structured data" links in a Web page header that links the content to published concept definitions such as those of schema.org. This provides more rigor to the interpretation of such content and the Web API's through which it may be accessed. More particularly, this structured data enhancement may be leveraged by Web search engines to better "understand" content and apply it to answer questions posed by Web search queries directly.

A separate engineering report, the Disaster Pilot 2021 Structured Data Engineering Report (OGC Document 21-054) covers this particular activity of the Pilot in much greater detail.

5
COMPONENTS

5 COMPONENTS

Each of the following sections describe the contributions, results, and lessons learned for each Pilot participant

1. Goals of participation

2. Description of the contributed component(s) and role in the Pilot architecture

3. Technical design and implementation

4. Scenario workflows and recipes

5. Technology interchange experiments

6. Challenges and lessons learned

7. Next steps: updates and applications

The documented contributions are shown in the following wiring diagram by number, type, and role:
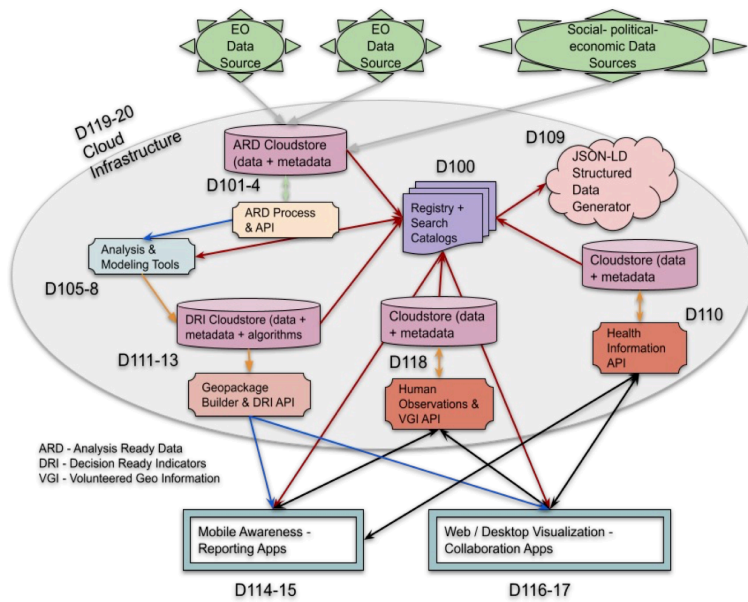
Figure 8 — Wiring diagram for dp21 information ecosystem components.

The goals for many of these components were to address specific challenges to information flow or pivotal points of interoperability between collection of EO data and delivery of usable decision guidance to practitioners on the ground or to the public.
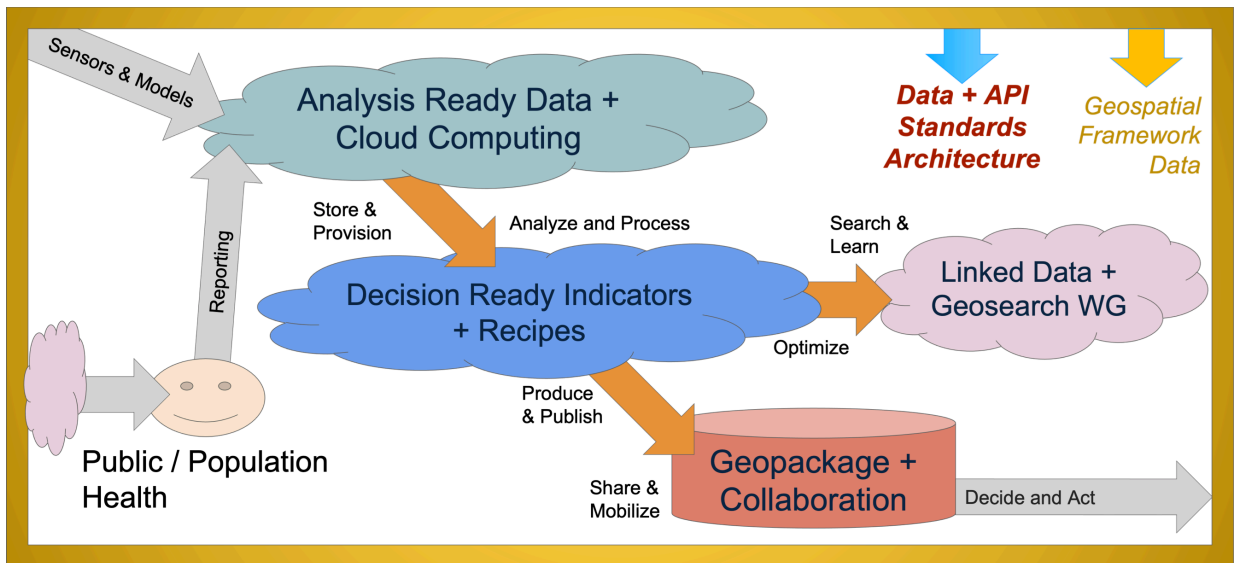
Figure 9 — Data flow diagram and pivotal points of information interoperability for dp21 information ecosystem components.

5.1. Health SDI Component (GeoSolutions)

5.1.1. Introduction

GeoSolutions has contributed to the Disaster Pilot by developing and deploying the platform used and operated by HSR.health and by other participants including Safe Software, RSS Hydro, and Skymantics, relying on the GeoNode open source project for which GeoSolutions is the main contributor.

The platform has played the core role of a Health SDI by the following.

- Providing the ability to ingest, store, disseminate, and make discoverable, ARD and DRI products, health indicators, and risk indices to other Health SDIs (e.g., Canadian Geospatial Platform, USGS GeoPlatform, and the AmeriGEO GeoPlatform, and platforms provided by other participants, such as StormCenter), the general public, and emergency response and public health decision makers through the OGC API, GeoPackage, WMS, WFS, WPS, and other OGC and open data standards.

- Supporting pen data standards to ingest, process, store, disseminate, and visualize raw data, ARD, and DRI from fellow participants including Safe Software, RSS Hydro, and Skymantic, as well as the collection and ingestion of on-the-ground data from field personnel, including but not limited to: health data, location and availability of existing healthcare facilities and resources, transportation information, road conditions, location of temporary or field health facilities and resources, environmental conditions, and other data relevant to situational awareness.

- Simplifying capability for end users to research on the web information about disasters by looking it up in search engines. To allow this, it is critical that those publishing the data and those reading the data have a common understanding of the domain, which can be achieved through well-defined semantics that describe precisely how the published data should be interpreted without ambiguities. Geosolutions' approach to this task has been to publish structured geospatial data using the JSON-LD format through a GeoServer OGC API implementation. The JSON-LD context was defined based on schema.org vocabulary.

5.1.2. Description of the contributed component(s) and role in the Pilot architecture

This section provides a description of the GeoNode platform and of the role it played in the overall architecture while the next section provides additional information on the technical implementation for the Disaster Pilot.

GeoNode is an open source project that was originally developed under the guidance of a number of organizations led by the World Bank *Global Facility for Disaster Reduction and Recovery (GFDRR)* group more than 15 years ago and that has since then been deployed all over the world to help users open up existing geospatial data to the web. Organizations like World Bank, UNMISS, KRIHS, UN WFP, UN IFAD, UNESCO, UN FAO, UNEP, and others rely on GeoNode to support various projects across the globe. GeoNode is available for free through a GPL license.

GeoNode provides all the functionalities to create a complete **GIS platform** as it provides data and metadata management as well as rapid mapping capabilities to go from raw data to maps, dashboard, and geostories with no development. With GeoNode, users can easily set-up a geospatial infrastructure to enable users to ingest, share, browse, preview, visualize, and disseminate data

GeoNode allows end users to upload assets (videos, documents, sound files, and more) as well as geospatial data (shapefiles, csv files, and geotiff files) both via a user-friendly interface as well as programmatically with the potential to define fine grain permission with respect of users and level of access (e.g., we can have private resources, public resources, or we can decide that certain resources are visible to anyone but only certain groups of users can download them).
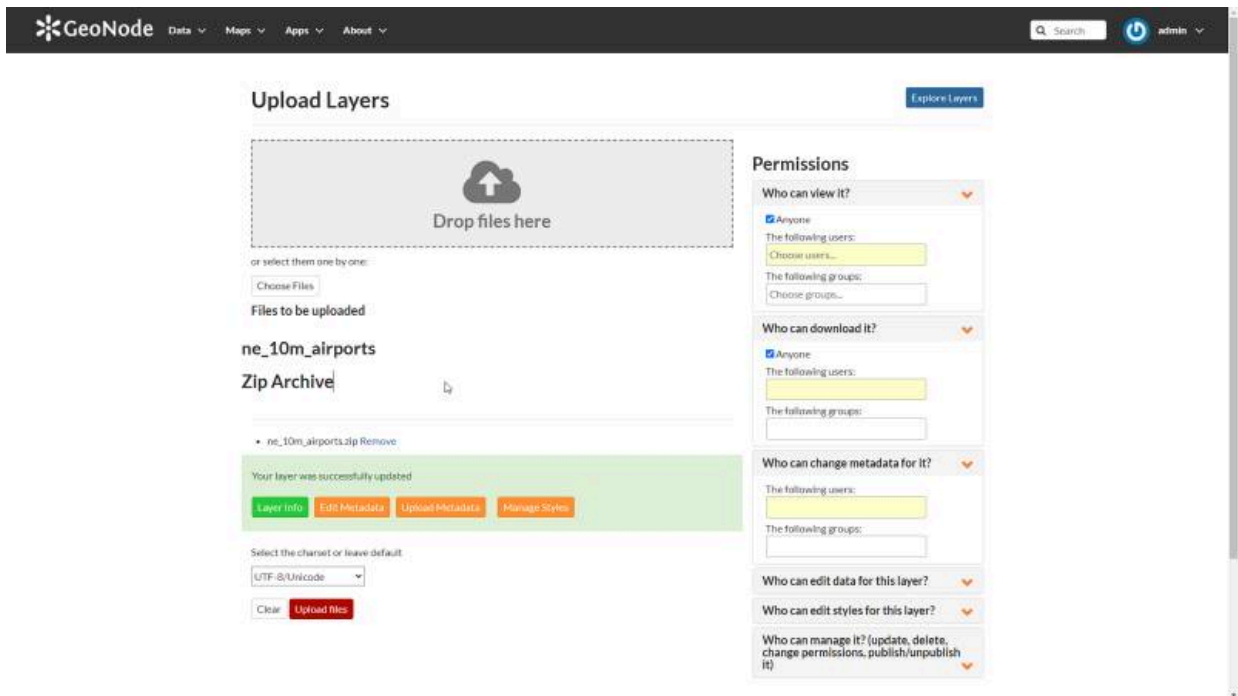
Figure 10 — Uploading a shapefile in GeoNode through
the user interface (with permissions management).

GeoNode provides an easy-to-use visual styler which can be used to create compelling
visualizations for geospatial data using OGC SLD standards as well as GeoServer own GeoCSS
styling language.
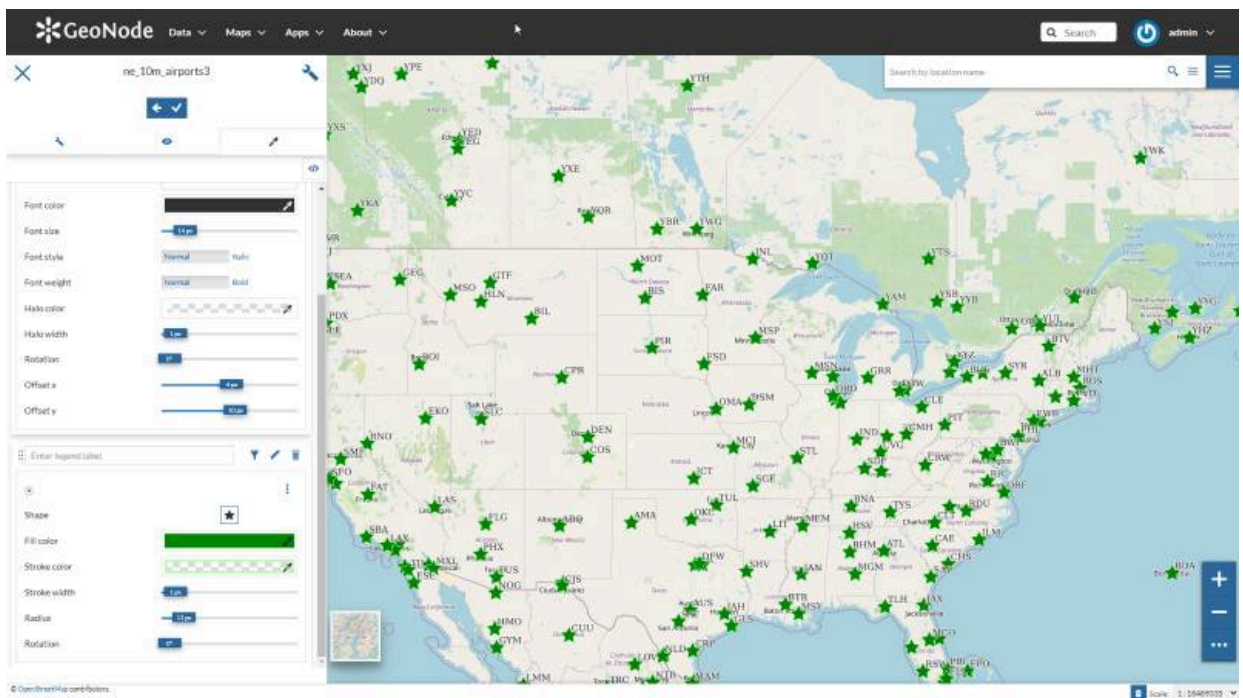
Figure 11 — Styling vector data in GeoNode through its integrated visual styler.

ISO Metadata management is fully supported in GeoNode in order to make data discoverable both by using GeoNode integrated portal as well as by external software thanks to GeoNode OGC CSW API.

Assets and geospatial data can be then discovered through GeoNode geoportal and accessed, edited, downloaded (as GeoPackage files for instance), and visualized with advanced charting capabilities through its integrated webgis based on the open source MapStore product.
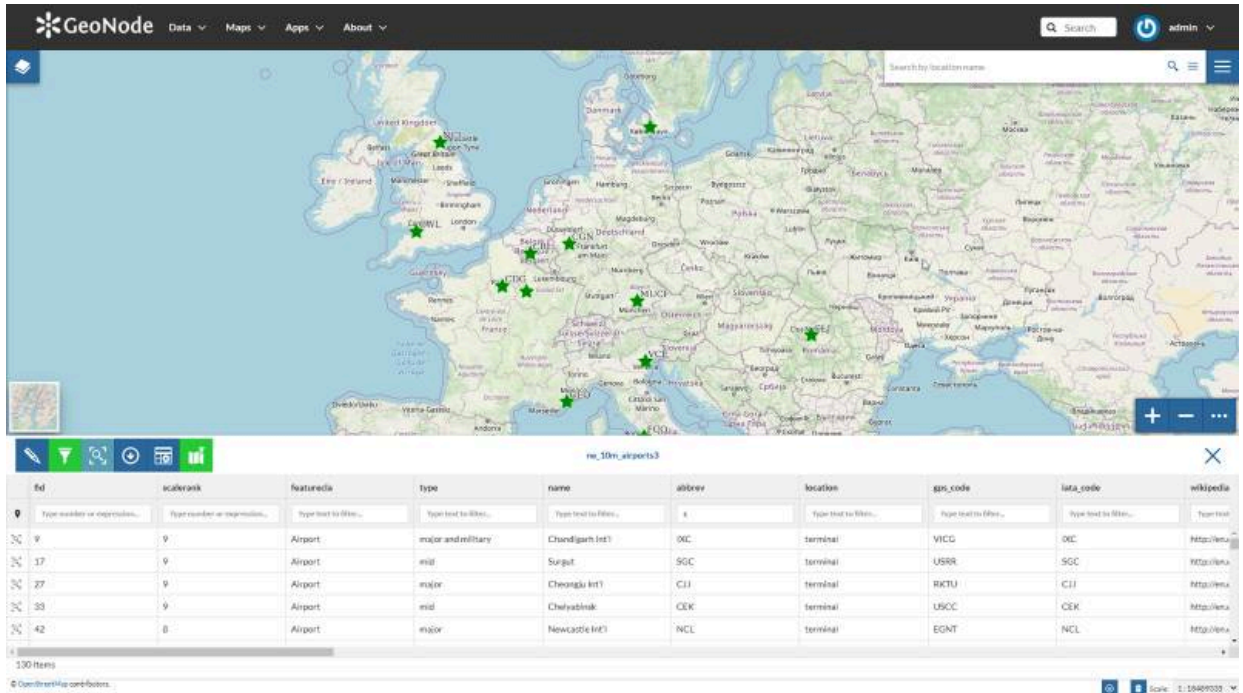
Figure 12 — Discovery of geospatial data in GeoNode.
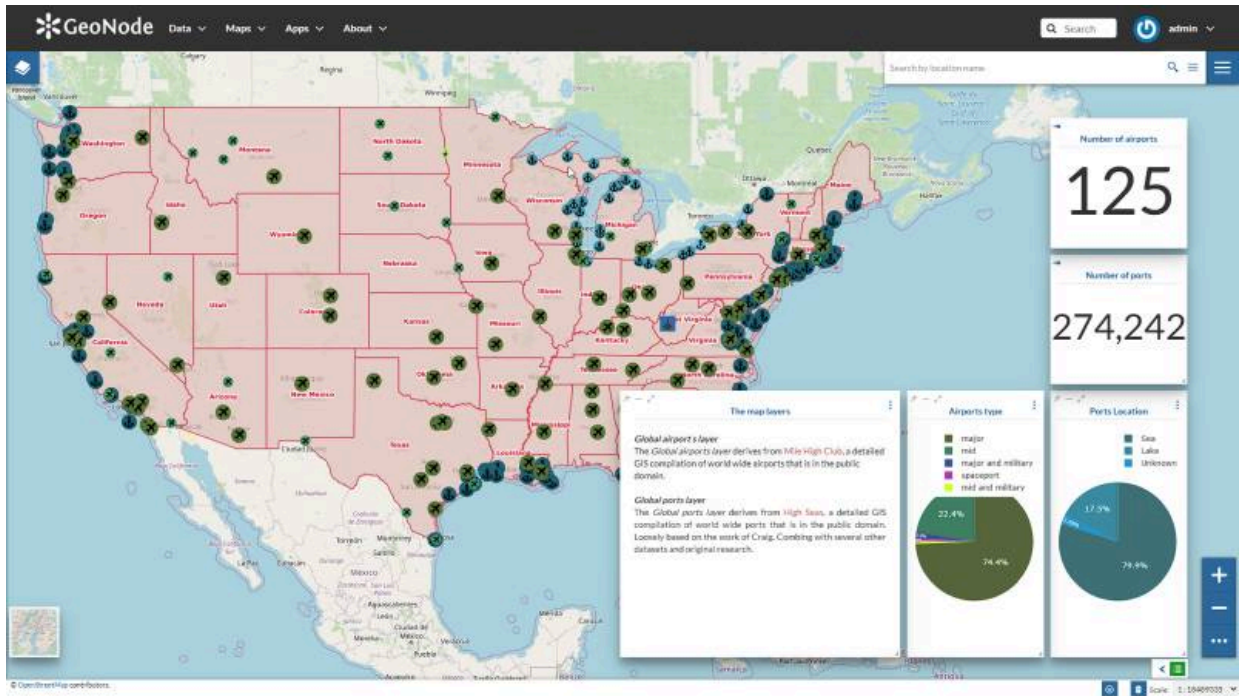
Figure 13 — Discovery of geospatial data in GeoNode.
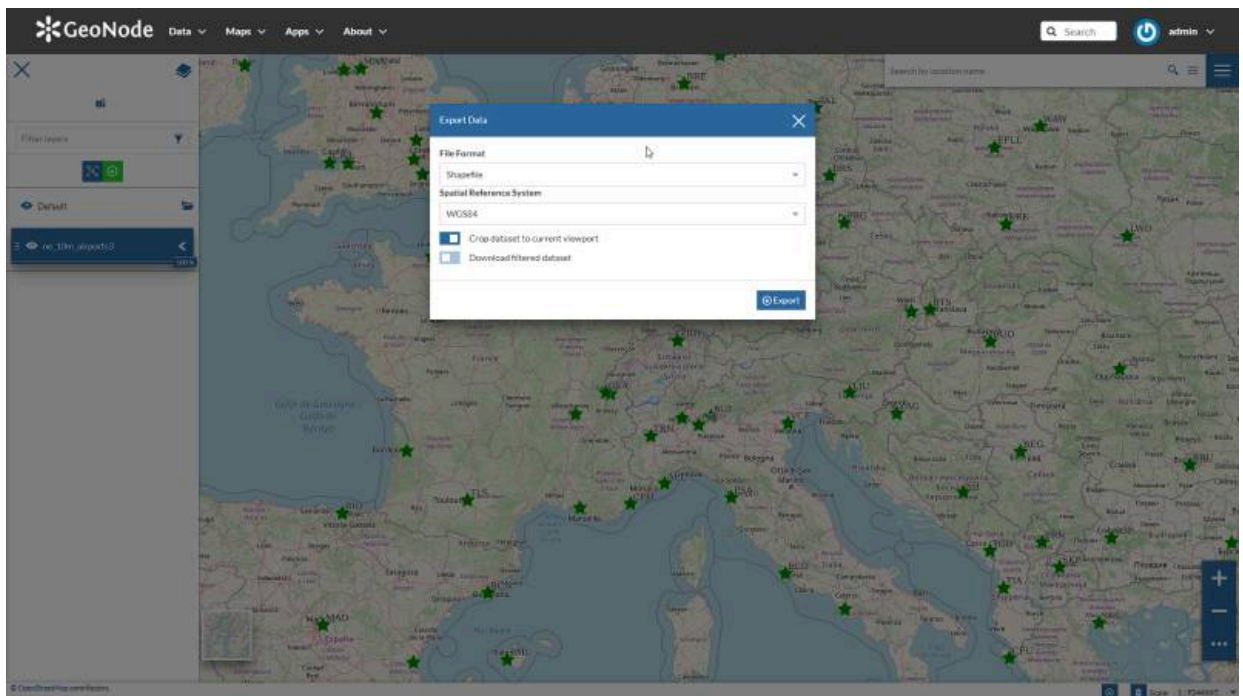
Figure 14 — Discovery of geospatial data in GeoNode.

More sophisticated visualization capabilities are also supported like temporal series, geostories for storytelling and dashboards.
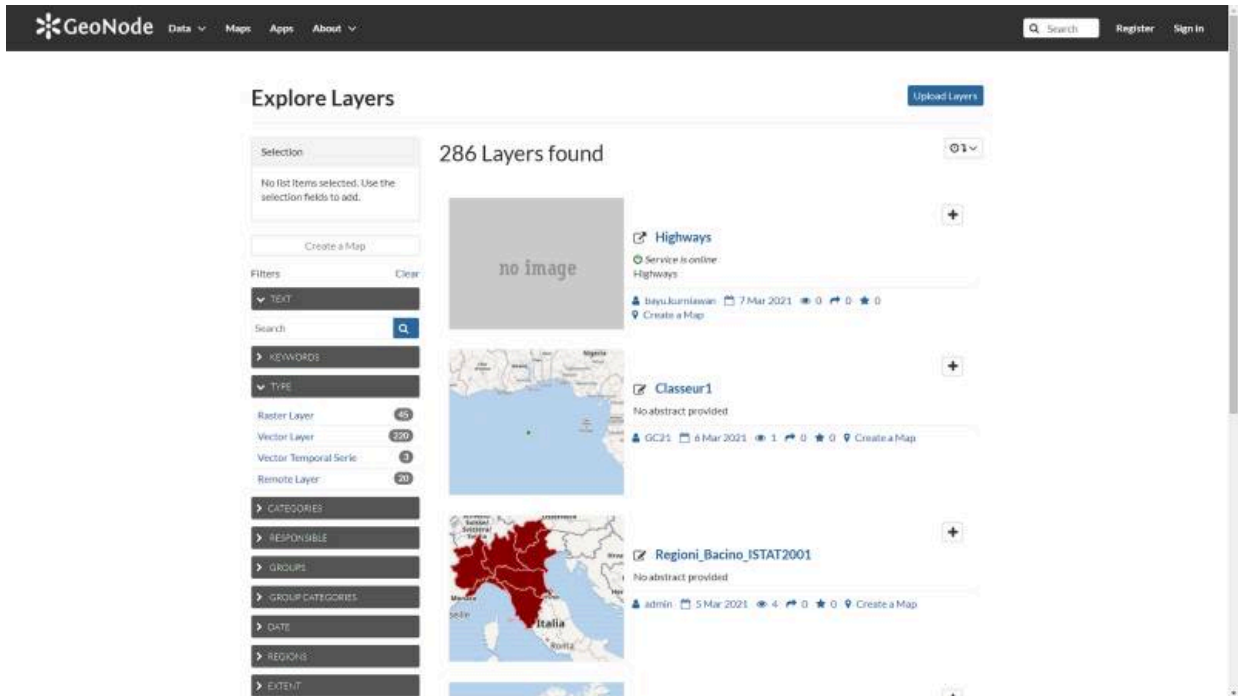
Figure 15 — Discovery of geospatial data in GeoNode.
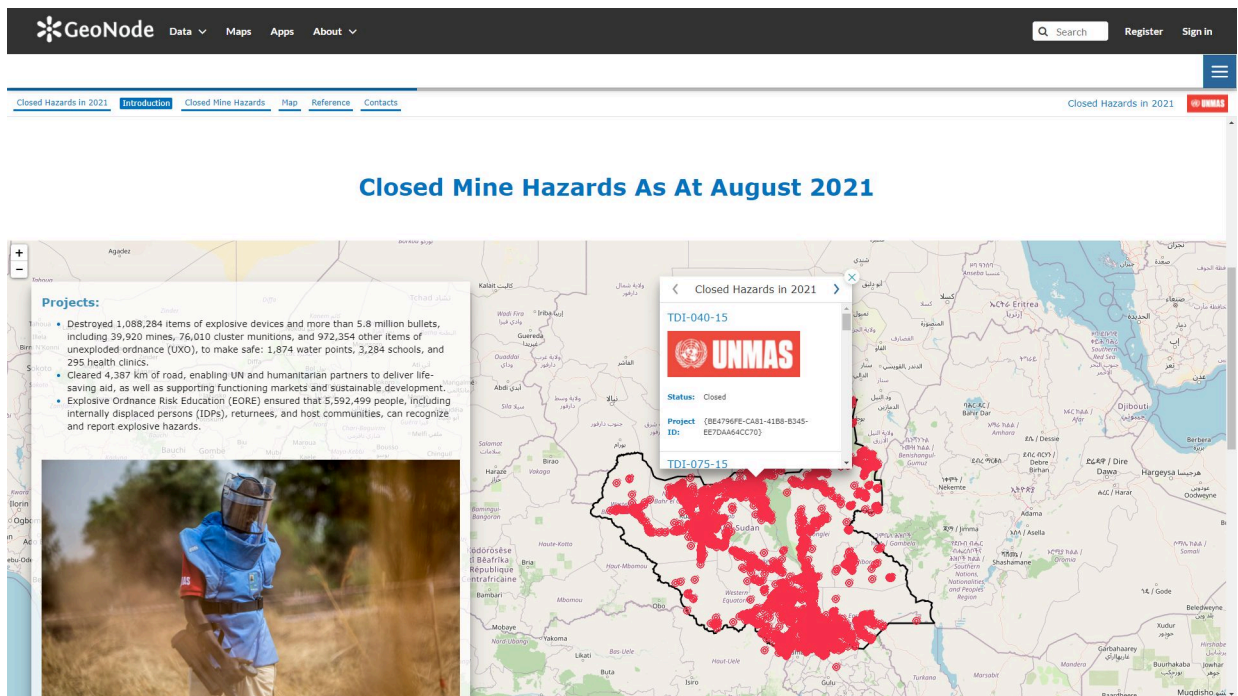
Figure 16 — GeoNode geostory fron UNMISS in action.
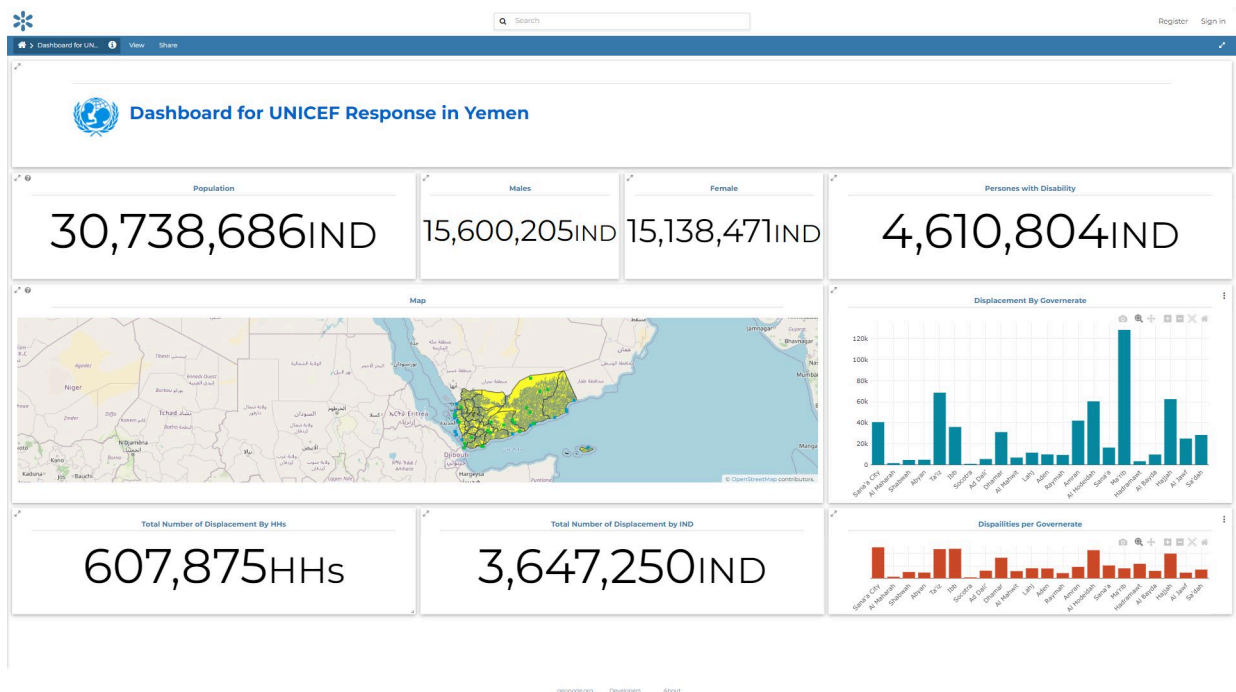
Figure 17 — GeoNode dashboard from UNICEF in action.

5.1.3. Technical implementation

GeoNode is composed by a suite of well-known enterprise products, the main ones are as follows.

- GeoServer, the leading open source geospatial server.

- MapStore, which provides webgis capabilities for GeoNode users.

- PostGIS, which is used to ingest and store vector data, as well as metadata and own GeoNode model.

- Django, the leading Python web framework which is used to implement GeoNode portal, GeoNode REST API, and GeoNode administration functionalities.

- pyCSW, a leading open source implementation of the OGC CSW protocol for the dissemination of ISO metadata. It is integrated in GeoNode to provide OGC CSW catalogue services to external applications.

5.1.4. Health SDI Implementation

The GeoNode product was customized to support HSR (and other partners') work on the Health SDI through a deployment in AWS (this is discussed in the HSR section). Particular effort was

made on the support for Structured Data via the Linked Data support from GeoServer. GeoNode provided support for a variety of functionalities and scenarios.

- **D100.1 Data Discovery Component**, as it has been used by partners to upload data for the various scenarios to make it available to various stakeholders via its portal as well as its OGC CSW API.

- **D109.1 Structured Web Data Component: GeoSolutions**, through the use of <u>GeoServer JSON-LD</u> extension which has been used to publish structured health related data.

- **D117.2 Analysis Component**, since GeoNode provides a web client application due to its MapStore integration which is able to discover, request, and download both ARD and DRI products for analysis and visualization by analysts and decision makers. The component shall also support virtual collaboration between distributed stakeholders both in the interpretation of such datasets, and in the design and selection of "recipes" for generation of decision-ready information products.

5.2. Mobile Awareness Reporting Apps | Web/Desktop Visualization Component (GISMO)

- Authors: Theo Goetemann, Jiin Wen, Amy Jeu

5.2.1. Introduction

GISMO, in collaboration with Ella Surveys and Basil Labs, has contributed to the 2021 Disaster Pilot through a voice surveying solution whose goal is to provide live, on-the-ground information, crowd-sourced from residents and first responders. Though this system is not meant to replace 911, cases where 911 is overwhelmed are exactly the situations where the ability to handle large scale information from the public are most critical. Additionally, prior to or following a crisis, the ability to gauge community sentiment as well as monitor on-the-ground information and adapt and iterate on the questions asked and analyzed via voice surveys are similarly important features. Through this voice survey platform, our team provides an environment where those in the offices can quickly discern and prioritize where to focus efforts and resources based on incoming data.

For the 2021 Disaster Pilot, the Mobile Awareness-Reporting Apps and Web/Desktop Visualization Collaboration Apps components were addressed. The Pilot focused on four essential elements.

- Scalable exploitation of earth observation data on hybrid cloud platforms and other technologies.

- Visualization of the right information at the right time using GeoPackage data containers and GeoPackage viewers.

- Generation of GeoPackage offline containers to allow taking all relevant information into the field.

- Capability for "structured data" to enable web search engines to push disaster-relevant information up in search results during an ongoing disaster.

5.2.2. Description of the contributed component(s) and role in the Pilot architecture

The figure below shows where GISMO fits into the DP21 Pilot. At a fundamental level, this solution permits surveyors access to both raw (voice responses in the original language as well as English) and analysis-ready data (already processed and categorized based on keywords into topic bins) which can then be mapped and visualized. Questions can be edited after survey creation, meaning as a crisis evolves, surveyors have the ability to continuously engage with relevant questions and gain up-to-date feedback.

In practice, this voice survey solution can touch on multiple departments and roles, from emergency managers to first responders to city officials, GIS specialists, and more. The key element to note is that the barriers between these roles are often barriers and information silos, particularly given an evolving crisis scenario. The on-the-ground information which can be classified and mapped in real-time can serve as an equalizing force so that every stakeholder is on the same page. These stakeholders can include peer organizations in the Pilot. The following figure illustrates this flow of information collection and feedback.

5.2.3. Technical design and implementation

This solution is fundamentally quite simple. The process is as follows:

- a surveyor creates a survey and deploys it by sending out a URL or QR codes;

- respondents speak to their phones;

- voice is transcribed to text;

- text is translated to English as need be;

- using easy drag-and-drop functionality, surveyors can assign keywords to topic bins;

- topics are assigned to each response; and

- since surveys ask for the mobile device's geolocation, responses can be imported via CSV, Rest API, or GeoRSS into mapping and visualization platforms.

The following illustrates the solution's flow.

Once responses have been classified, individuals who wish to visualize responses geographically can use GeoRSS to import the data, which, depending on the platform, can be set to regularly update itself, thus providing real-time visualizations.

The following is the API schema for returned data.

```
// API SCHEMA FOR ELLA

// Axios request example

return new Promise((res, rej) => {
        axios({
            method: "POST",
            url: serverUrl + '/getAllAnswersForSurvey',
            data: {
                apiToken: <TOKEN YOU'RE PROVIDED WITH>,
    surveyUid: <YOUR SURVEY UID> (get this from url bar when you open the
survey)
            },
        }).then(response => {
            if (response.status === 200) {
                res(response.data)
            }
        })
            .catch(function (error) {
                rej("Something went wrong, please try again!")
            });
    })


// SAMPLE RESPONSE JSON
{
    "questions": [
        {
            "orderNumber": 1,
            "question": "How are you?",
            "surveyUid": "d27d087-3d5a-55fc-4a2e-347c251d38db",
            "type": "text",
            "collectionName": "new collection",
            "questionUid": "6da503b-3deb-00f-523a-010187ad5e7a",
            "surveyName": "Business Survey 2021",
            "translations": [],
            "collectionUid": "4ca2f76-0787-a01-6d-df1cabb8b888",
            "userUid": "UHcoxvRQI8gbYPJeg1VOvmUYs2Ct"
        }
    ],
    "answers": [
        {
            "answerUid": "01750dd-5826-d443-62f-c05dadf042a",
            "question": "How are you?",
            "surveyUid": "d27d087-3d5a-55fc-4a2e-347c251d38db",
            "gpsLocation": "{}",
            "questionUid": "6da503b-3deb-00f-523a-010187ad5e7a",
            "answer": "test",
            "languageObject": {
                "Name": "English (United States)",
                "BCP-47": "en-US"
            },
            "respondentUid": 1634545784760,
            "collectionUid": "4ca2f76-0787-a01-6d-df1cabb8b888",
            "userUid": "UHcoxvRQI8gbYPJeg1VOvmUYs2Ct",
            "timestamp": 1634545792406
        },
        {
            "answerUid": "018a0d4-a815-f6ba-027a-f13bf3a607",
            "question": "How are you?",
            "surveyUid": "d27d087-3d5a-55fc-4a2e-347c251d38db",
```

```
            "gpsLocation": {
                "altitude": null,
                "heading": null,
                "latitude": 37.09024,
                "accuracy": 1489640.756336928,
                "altitudeAccuracy": null,
                "speed": null,
                "longitude": -95.712891
            },
            "questionUid": "6da503b-3deb-00f-523a-010187ad5e7a",
            "answer": "clear",
            "languageObject": {
                "Name": "English (United States)",
                "BCP-47": "en-US"
            },
            "respondentUid": 1634546983921,
            "collectionUid": "4ca2f76-0787-a01-6d-df1cabb8b888",
            "userUid": "UHcoxvRQI8gbYPJeg1VOvmUYs2Ct",
            "timestamp": 1634546995596
        },
        {
            "answerUid": "1dee214-25ea-b115-0d4b-d6503fbed5e",
            "question": "How are you?",
            "surveyUid": "d27d087-3d5a-55fc-4a2e-347c251d38db",
            "gpsLocation": {
                "altitude": null,
                "heading": null,
                "latitude": 44.8180773,
                "accuracy": 14.39,
                "altitudeAccuracy": null,
                "speed": null,
                "gpsLocation": {},
                "longitude": 20.4241145
            },
            "questionUid": "6da503b-3deb-00f-523a-010187ad5e7a",
            "answer": "test",
            "languageObject": {
                "Name": "English (United States)",
                "BCP-47": "en-US"
            },
            "respondentUid": 1634546843135,
            "collectionUid": "4ca2f76-0787-a01-6d-df1cabb8b888",
            "userUid": "UHcoxvRQI8gbYPJeg1VOvmUYs2Ct",
            "timestamp": 1634546880110
        }
    ]
}
```

Figure 18

5.2.4. Scenario workflows and recipes

One of the "game-changer" elements of putting voice response first in surveying is the new ability to freely ask open-ended questions without the worry of a low response rate. The following are some examples of open-ended questions. As you are reading them, imagine what the multiple-choice equivalent would be and the quality of feedback each question would receive.

- "What challenges are you currently facing?"

- "What supplies do you need?"

- "How has seasonal flooding impacted your decision to stay or sell and move elsewhere?"

- "Tell us what's happening in your area. How have conditions changed?"

As such, one of the main components of this project was to establish the right kinds of questions to ask in the case of flooding in Canada or Peru, blending both multiple-choice, short answer and open response. Through conversations with other Pilot participating organizations such as HSR, among others, our team developed the following questions.

- Do you need to be rescued?

- How high is the water?

- How fast is the water flowing?

- Tell us about the street conditions.

- Do you have one of the following medical conditions?

- Do you have medication for the next three days?

- If you are injured, please tell us the severity of the injury and your condition.

- Do you have power and water?

- What supplies do you need?

5.2.5. Challenges and lessons learned

5.2.5.1. Voice responses are incredibly powerful when leveraged appropriately, but a generation of scientists will need to shift their mindsets towards open-response.

It truly has been a learning experience to wrap our heads around the kinds of questions that can be asked once the barriers to inputting information and analyzing responses have been lifted. Having conversations with professionals who regularly leverage surveys in their daily work and hearing their own thoughts on the new doors that open up when you can ask open ended questions was inspiring and something to continue exploring, as it is conceptually, a foreign concept for many researchers, who are more familiar with manual methods of narrative analysis and topic coding for open responses.

5.2.5.2. Engagement of end users during pilot projects

While technology can deliver incredible results, for a survey platform, those results are very much grounded in the communities that are to be surveyed, and having closer access to officials

and leaders in those communities, particularly in Peru, would continue to strengthen both the solution technically as well as add nuance and know-how to the questions we choose to ask in the surveys.

5.2.5.3. The key to cracking open-response survey analysis is efficient, easy-to-use topic categorization functions.

There is always a balance between usability and sophistication, and in the case of topic categorization, usability is paramount. We built a drag-and-drop interface for classification of topics and corresponding keywords not only for the technical ease of use, but also to assist users conceptually. Rather than bring in semi-supervised learns and rapid training dataset creation (for surveys that many of the times will have very little data to work with due to smaller response samples), keywords corresponding to topic bins was an approach that users felt comfortable with letting the machine take over in assigning topics to responses.

5.2.6. Next steps: updates and applications

During the Pilot, we were able to engage a geo-located, multi-lingual survey platform, which will be accessible here for use: https://app.ellaxyz.com/.

There are a number of other solutions to build out in the same vein as Ella Surveys, which all revolve around the singular idea of technology removing barriers between communities and decision makers.

5.3. Spatial Data infrastructure & Reporting Component (HSR.health)

____

5.3.1. Goals of participation

HSR.health demonstrated a Health SDI that currently supports COVID-19 response globally and is aligned with the draft findings of the OGC Health SDI CDS. Further the pilot effort demonstrated the following.

- The potential impact of natural disasters on pandemic response needs.

- The curation or creation of datasets as necessary for pandemic and co-incident natural disaster response.

- The discoverability and availability of relevant datasets; effective communication and sharing of ARD, DRI, health indicators; and risk indices through a Health SDI, and how these datasets can be leveraged to help mitigate the cascading strain from co-incident pandemic and natural disasters.

- The ability to maintain pandemic response during natural disasters, which was demonstrated for landslide and flood scenarios and applicable to natural disasters broadly.

- The collection and ingestion of on-the-ground data from field personnel, including but not limited to:

  - health data;

  - location and availability of existing healthcare facilities and resources;

  - road conditions;

  - location of temporary or field health facilities and resources;

  - environmental conditions; and

  - other data relevant to situational awareness.

- Our Health SDI follows open data standards and will be able to ingest, process, store, and analyze raw data, ARD, and DRI from fellow participants including Safe Software, RSS Hydro, and Skymantics.

- The ability to share, export, communicate, and make discoverable, ARD and DRI products, health indicators and risk indices from our Health SDI to other Health SDIs (e.g., Canadian Geospatial Platform, USGS GeoPlatform, the AmeriGEO GeoPlatform, and platforms provided by other participants, such as StormCenter), the general public, and emergency response and public health decision makers through the OGC API, GeoPackage, WMS, WFS, WPS, and other OGC and open data standards.

- The ability to support environments of widely variable connectivity. Our Health SDI as well as the Medical Supply Needs Indices aids the overall emergency response from manufacturer-to-supplier-to-end-user and can be adapted to leverage standard identifiers such as GS1 GLN identification keys.

5.3.2. Description of the contributed component(s) and role in the Pilot architecture

The primary purpose of the contributed components was to provide health related data and analytics to the pilot as well as expand the existing health spatial data infrastructure data model to aid in data sharing within the pilot. The following components were demonstrated in the pilot.

- A Health-focused Spatial Data infrastructure (SDI) that follows open data standards and will be able to ingest, process, store, and analyze raw data, ARD, and DRI from fellow participants including Safe Software, RSS Hydro, Skymantics, and others. With the ability to share, export, communicate, and make discoverable, ARD and DRI products, health indicators and risk indices from our Health SDI to other Health SDIs (e.g., Canadian Geospatial Platform, USGS GeoPlatform, the AmeriGEO GeoPlatform, and platforms provided by other participants, such as StormCenter), the general public, and emergency

response and public health decision makers through the OGC API, GeoPackage, WMS, WFS, WPS, and other OGC and open data standards.

- Medical Supply Needs index, Transmission Risk Index, Mortality Risk Index, and hospital locations for Peru and Louisiana.

- Documentation and research on health and pandemic data sources for expanding upon the data model advanced in the OGC's Health Spatial Data Infrastructure Concept Development Study.

5.3.3. Technical design and implementation

5.3.3.1. Technical Design

The component data of the Risk indices were retrieved and processed through python scripts. After initial processing the data was aggregated together to create each index. The indices were then stored in our postgres database which is used as a datastore for the Health Spatial Data Infrastructure. Views of the data were published from the datastore to the geonode and geoserver components of the SDI which allow them to be searchable and downloaded through OGC open standards including WMS, WFS, and as a geopackage. Additionally, in-depth dashboard analytics were provided through Tableau Online dashboards that pull data from the postgres database offering a large amount of flexibility for filtering and analyzing the data without the need to download the data locally. The following image shows the infrastructure flow diagram.

5.3.3.2. Implementation

The Health SDI was built and hosted on AWS for access by interested pilot participants. After the creation of the risk indices and data were uploaded to the Health SDI, a Tableau Online dashboard was established to enable additional analytics and viewing of the data. The findings from the research into the granularity and availability of data were shared during multiple presentations throughout the pilot and were also covered in the challenges and lessons learned section of this document.
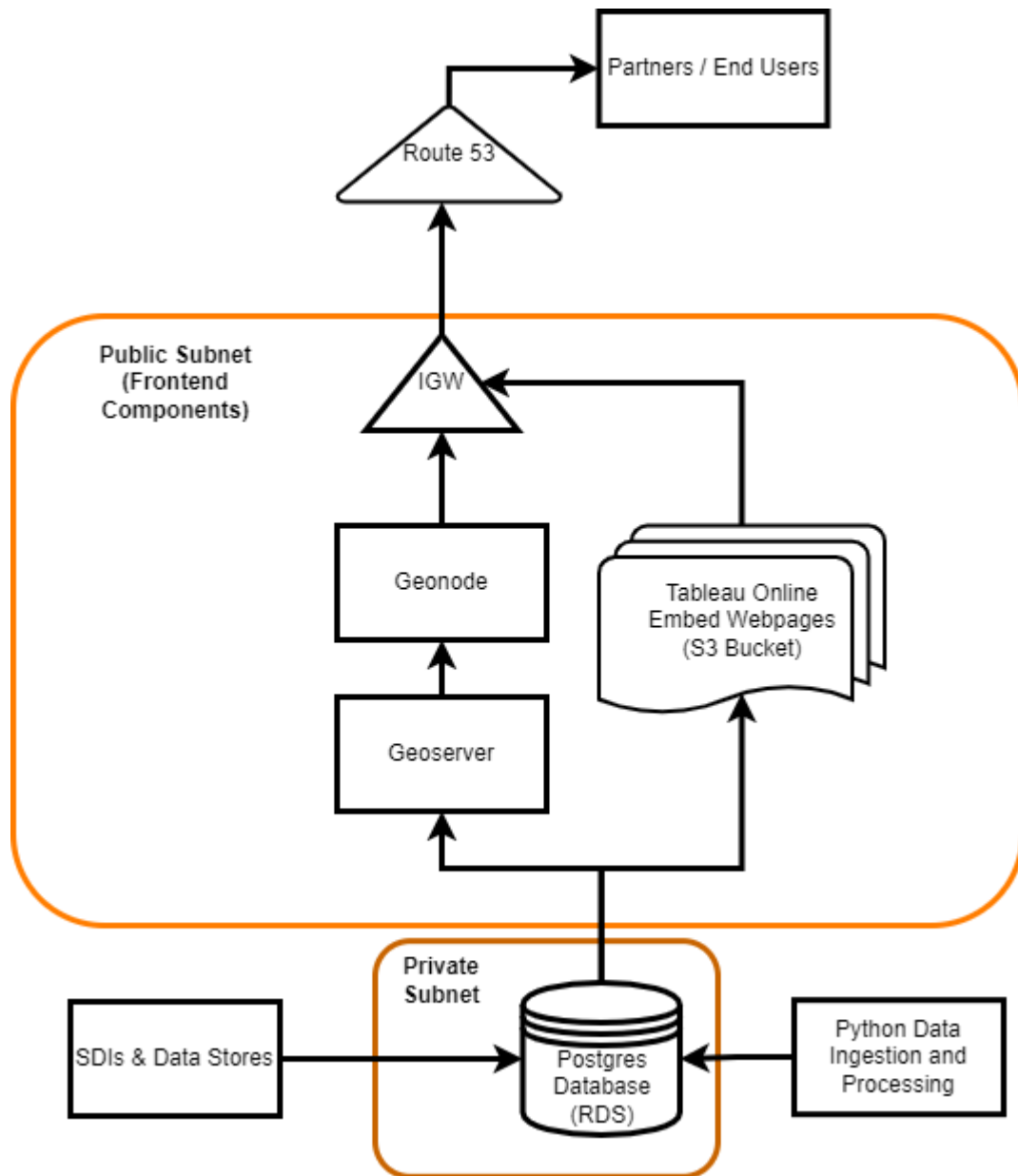
Figure 19 — HSR AWS Infrastructure Diagram for the Health Spatial Data Infrastructure

5.3.4. Scenario workflows and recipes

HSR.health calculated, produced, and shared the Medical Supply Needs Index for both the US and Peru Scenarios during the 2021 Disasters Pilot. The purpose of the Medical Supply Needs Index is to estimate medical supply need based upon the current spread of the COVID-19 pandemic. It can also be modified to incorporate additional supply needs and routing based upon a coincident natural disaster. One of the components of the Medical Supply Needs Index is the Pandemic Risk Index, which is comprises both our Mortality and Transmission Risk Indices.

The Mortality Risk Index utilizes data on population demographics and the prevalence of comorbidities to identify the risk to the underlying population of severe illness, hospitalizations, or mortality due to the COVID-19 pandemic.

The Transmission Risk Index utilizes data on population, case counts, area, and human mobility to identify the spread of COVID-19 by geographic area. The Pandemic Risk Index combines these two indices to represent both the spread of the pandemic and the health risk that the pandemic poses.

The Mortality, Transmission, and Pandemic Risk Indices are normalized to fall between 0 and 100. Current generalized recommendations from those indices are:

- 0-25 is low risk;

- 25-75 is moderate risk; and

- 75-100 is high risk.
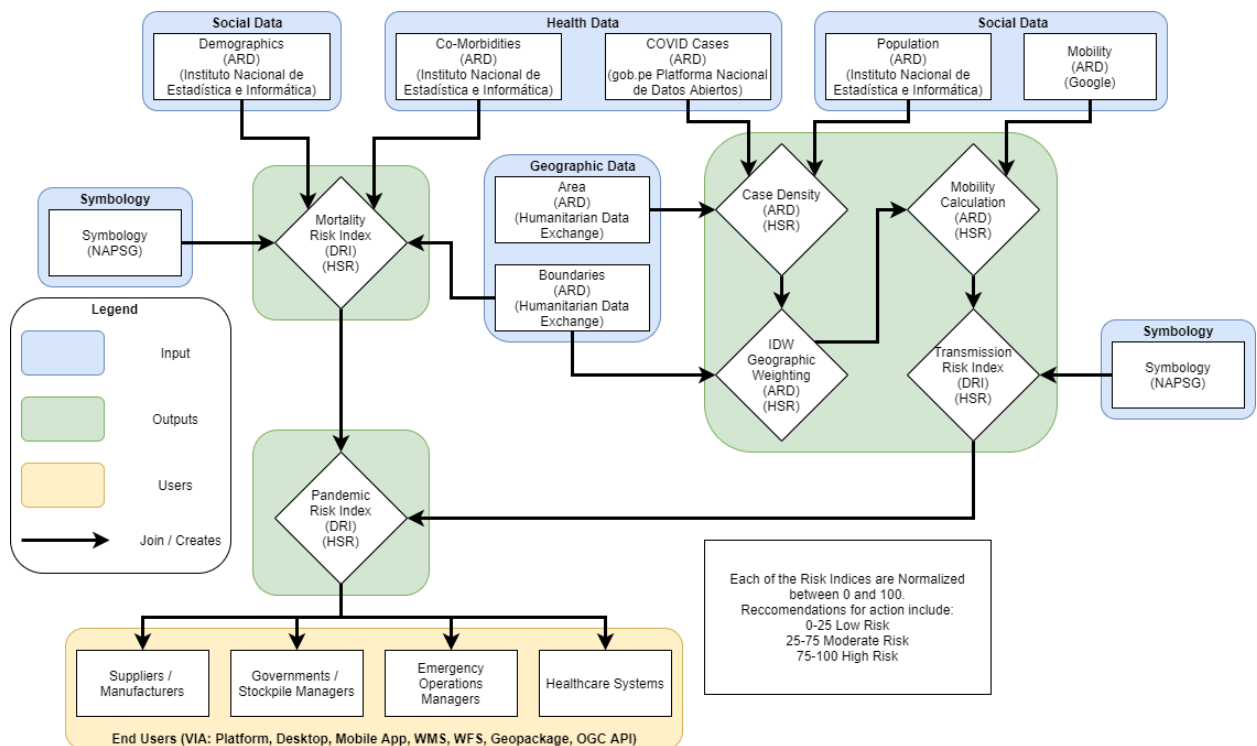
Figure 20 — Peru Transmission, Mortality, and Pandemic Risk Index Workflow Diagram

The Medical Supply Needs Index produces three levels of need: high, low, and index-based levels of need. The high and low levels of medical supply needs are based on hospitalizations, ICU hospitalizations, supply utilization by type (in this case PPE such as gowns, gloves, and masks), the number of healthcare workers, and the number of first responders like police, firefighters, and EMS. Once the high and low estimates are calculated, the Pandemic Risk Index is incorporated to produce the supply level based upon the spread of the pandemic and the health risk to the underlying population. Depending on the individual hospital's risk tolerance for medical supplies, they can decide to use any of the high, low, or index based need levels.

The recommendation for action to the decision makers is based on calculating the difference between the Medical Supply Needs Index, for the chosen risk level, and the current supply stock at each hospital and clinic, to determine when and how much supply is needed.
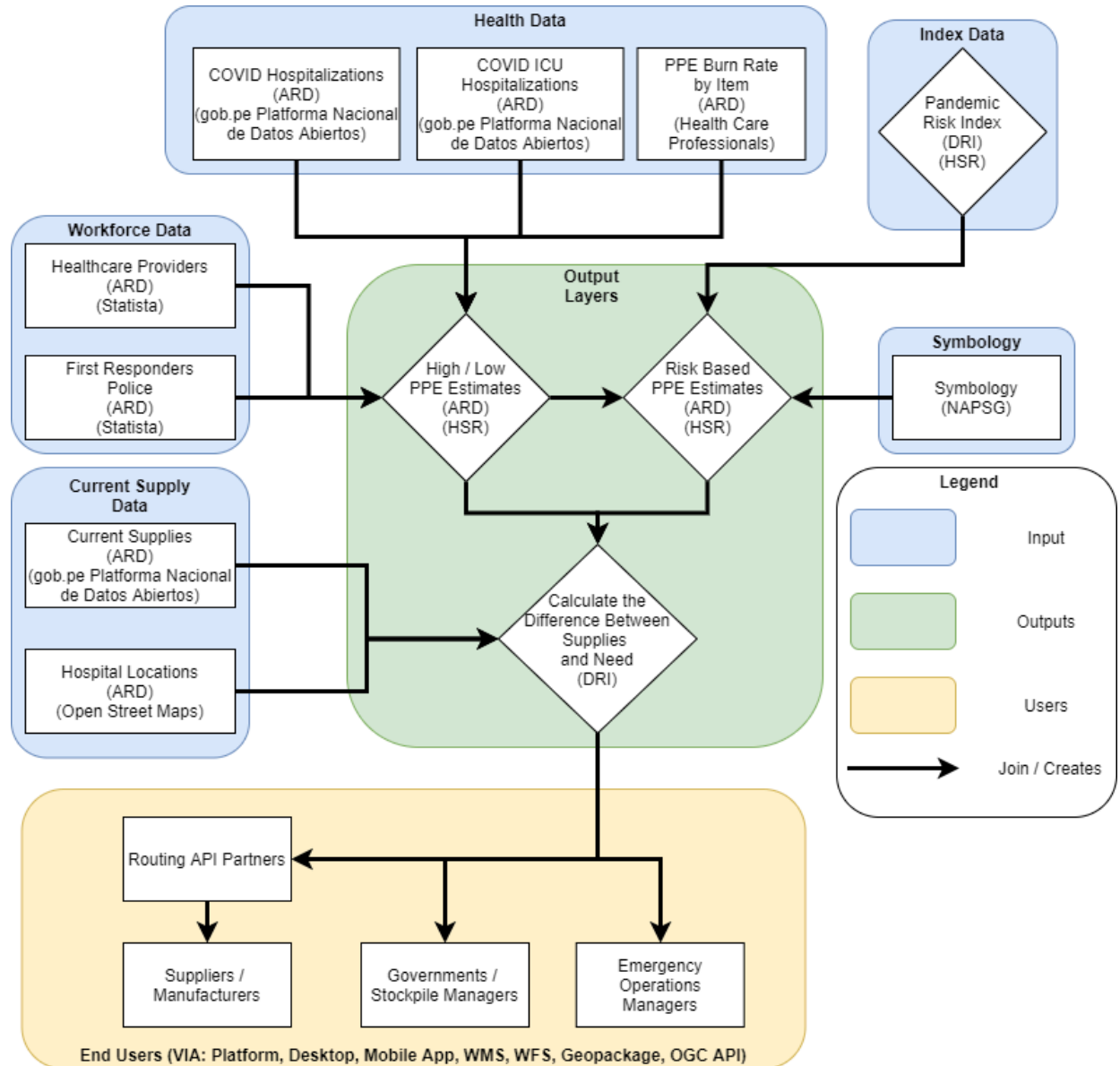
Figure 21 — Peru Medical Supply Needs Workflow Diagram

5.3.5. Technology interchange experiments

Data used by HSR.health, RSS Hydro, and Safe Software was uploaded and utilized on our Health SDI. The data hosted on our Health SDI was also shared with GeoCollaborate through Web Map Services (WMS). The shared data through WMS shown below are clinic locations in the Rimac River basin in Peru as well as a flood extent in the basin.

Figure 22 — RSS Hydro and Safe Software Layers on the Health Spatial Data Infrastructure

Figure 23 — HSR Data on Geocollaborate for Peru through WMS

5.3.6. Challenges and lessons learned

One of the contributions by HSR.health to the pilot effort was to expand upon the existing data model for the Health SDI CDS and identify the mismatch between ideal and currently available data.

The three main areas of data and indicators that HSR.health expanded upon include:

- Essential Foundation Data and Technology Data — which includes base layers such as road networks and hospital locations;

- Health Indicators for Disasters — such as identifying individuals with electricity-dependent medical devices and/or care, as well as infected water supplies; and

- Health Indicators for Pandemics — including information such as case counts, deaths, mortality risk, and vaccination rate.

In order to ascertain local health issues, the Census tract level of geography is the ideal granularity. The State of Louisiana makes available pandemic related information including case counts and deaths at the Census tract level which is one of the reasons why it was chosen for the US scenario. Numerous states do not make pandemic data available down to the Census tract or ZIP Code level. Data at the county or state levels are helpful, but limit the ability to perform local level analyses of health issues.

Establishing a centralized repository for health related data will help enable increased visibility of up-to-date health conditions and speed pandemic response and global public health decision making overall. Below is a full table of the expanded health and pandemic related data sources and their ideal and available temporal and spatial granularities for the state of Louisiana.

Table 1

Essential Foundation Data and Technology	LA Granularity	Target (Desired) Granularity	Temporal Granularity	Target (Desired) Temporal Granularity
Hazard and Vulnerability Areas (High Risk)		Affected Area		Daily
Imagery with x,y Coordinates or Grid addresses		Affected Area		Hourly
Address Database and GeoCoding Application		Point Locations		Annual
Habitation Layer (villages, homes, farms)		Point Locations		Annual
Demographic Layers	Census Block Group	Census Tract	Decennial, Annual	Monthly

Health System Boundaries (i.e., hospital catchment areas)	Zip Code	Census Tract	Annual	Annual
Health Facilities (e.g., hospitals, clinics, medical offices, health centers, pharmacies, labs, reserach facilities, medical schools, dental clinics, nursing homes, LTC, Rehab centers, etc.) and Resources Layers	Point Locations	Point Locations	One Time	Annual
Geographic Boundaries (e.g., States, Election Districts, ZIP Codes, Census Tracts)	Census Block Group	Census Tract	Decennial	Decennial
Emergency Dispatch Centers	Point Locations	Point Locations	Unknown	Annual
Social Facilities (See Code List in Health CDS)	Point Locations	Point Locations	Unknown	Annual
Diagnostic Testing Centers	Point Locations	Point Locations	One Time	Monthly
Critical Infrastructure Layers (E.g. water, power)		Point Locations		Annual
Social Vulnerability Index (Customized to Region)	Census Tract	Census Tract	Annual	Monthly
Vaccination Prevalence	Parish	Census Tract	Daily	Daily
Pre-Existing Conditions	Parish	Census Tract	Annual	Monthly
Social Resistence Index (SRI)		Census Tract		Monthly
Resistance to Modifying Behavior		Census Tract		Daily
Resistance to Vaccines (Vaccine Hesitancy)	Parish	Census Tract		Daily
Registry of Disabled, Device Dependent, Other Highly Vulnerable Individuals	Census Tract	Census Tract / Point Locations	Annual	Monthly
Resilient and Quickly Reparable Wireless Telecom Network		Point Locations		Daily
Road Network, Freight Train Routes	Poly Lines	Poly Lines	Up to Date	Annual
Helicopter and Aircraft Landing Zones	Point Locations	Point Locations	Up to Date	Annual
Inventory of Available Aircraft		Point Locations		Daily
Routes and Facilities for Critical Supply Chains	Poly Lines	Poly Lines		Daily

		Poly Lines / Point Locations	Daily
Tracking Shipments of Critical Supplies		Poly Lines / Point Locations	Daily
Mobile phone data (number of, utilization rate, location)		Point Locations	Real-time
Inventory and Location of Heavy Equipment		Point Locations	Real-time

Health Indicators for Pandemics	LA Granularity	Target (Desired) Granularity	Temporal Granularity	Target (Desired) Temporal Granularity
Pandemic tracking worldwide		Admin Level 3 / Census Tract	Daily	Daily
Symptomatic Travellers/Frequent Destinations		Point Locations		Daily
Infected Responders		Census Tract		Daily
Social Media Content Harvesting		Census Tract	constant	Real-time
Syndromic Surveillance (E.g. Drug sales, Health supply sales)		Census Tract / Point Locations		Daily
Internet Linked Thermometers (E.g. Kinsa)		Census Tract	Daily	Daily
Structured Voice & Written Messages		Census Tract / Point Locations	Real-time	Real-time
Structured Info from Political, Military, Religious, Community Leaders		County / School / Election Districts		Daily
Information from Health Care workers		Census Tract / Point Locations		Daily
Information from Health Care Facilities		Census Tract / Point Locations		Daily
Structured Info from 1st Responders		Census Tract / Point Locations		Daily

Information from Contact Tracers		Census Tract		Real-time
Wastewater Testing for Pathogen		Census Tract / Point Locations		Daily
Hotspot and Micro-Cluster Identification	Parish/Tract	Census Tract / Point Locations	daily	Daily
Testing and Diagnosis at First Contact with Health System		Point Locations		Daily
Diagnosis at Triage and After Hospitalization		Point Locations		Real-time
RO Measurements in Micro Areas		Census Tract / Point Locations		Daily
Predicted Increases in Illnesses		Census Tract / Point Locations		Daily
Inventory levels of medical devices, respirators, pharmaceuticals		Point Locations		Real-time
Utilization levels of medical devices		Census Tract / Point Locations		Real-time
Inventory levels of medical supplies (test kits, lab chemicals, syringes, sanitizers, alcohol swabs, cotton swabs, test tubes, masks, gowns, beds, oxygen, etc.)		Point Locations		Real-time
Raw materials for tests, vaccines, treatments		Census Tract / Point Locations		Real-time
Supply levels of Other Essential Supplies (E.g. food, water, sanitary products, etc.)		Census Tract / Point Locations		Real-time
Patient to Health Worker ratio	metro statistical areas, special areas inside, states	Census Tract / Point Locations	Quarter	Daily
Percent testing positive/negative	parish/tract	Census Tract / Point Locations	Daily	Daily
Positives by age, vulnerability, ethnicity, gender	parish/tract	Census Tract	Daily	Daily

Health Indicators for Disasters	LA Granularity	Target (Desired) Granularity	Temporal Granularity	Target (Desired) Temporal Granularity
Area and Population encompassed by Disaster and Cascading Effects	Tract	Census Tract	Hourly	Real-time
Area, Structures, and Population Covered by Floodwaters	Tract	Census Tract / Point Locations	Hourly	Real-time
Damaged and Destroyed Structures incl. Critical Infrastructure, Houses	Point	Point Locations		Real-time
Infected Responders		Census Tract		Daily
Homeless Population, Groups Seeking Shelter	Parish	Census Tract / Point Locations	Annual	Daily
Deaths Above Normal	Parish	Census Tract	Annual	Daily
Population of Power Outage Area	Tract	Census Tract		Real-time
Population with Compromised Water Systems	County	Census Tract	Quarterly	Real-time
Population in Area of Dangerous Air Pollution	Tract	Census Tract	Daily?	Real-time
Population in Area Lacking Communications	Tract	Census Tract	Biannual	Real-time
Incidents of Panic Buying and Looting		Census Tract		Daily
Weather Forecasts		Affected Area	Hourly	Hourly
Exposure (Cold, Heat)		Census Tract	Hourly	Hourly
Pandemic Spread	Tract	Census Tract	Daily	Daily
Respiratory Illnesses (e.g., Asthma)	Tract	Census Tract	Annual	Monthly
Digestive Illnesses		Census Tract		Monthly
Crushing Trauma		Census Tract		Daily
Drownings/Suffocation	State	Census Tract	Annual	Real-time
Mental Health	Tract	Census Tract	Annual	Monthly

Criminal Victimization	New Orleans/ Baton Rouge Only	Census Tract		Daily
Pathogen Identification In Water	Point	Point Locations	1-3 days	Real-time
Pathogen Identification in Vectors (e.g. Mosq)		Point Locations		Daily
Dangerous Chemicals in Water and Air	Tract	Census Tract	Daily?	Real-time

A challenge related to the pilot process is that for these written components it would have been have preferred to have gotten an outline of what was expected so that there would be more time during the length of the project to fill out the necessary sections as they were built.

5.3.7. Next steps: updates and applications

Future disaster pilot efforts can include the following.

- Expand upon the Louisiana Scenario to incorporate coincident natural disasters and the impacts of climate change.

- Explore applications of EO for healthcare and pandemic use cases.

- Expand the number of indices / analysis that we replicate in the Peru scenario:

  - Vaccine Needs Index;

  - Health Equity;

  - Variant Analysis / Prediction; and

  - Secondary Bacterial Index.

- Replicate our indices in Scenarios 2 (Red River, Canada) and 3 (Louisiana):

  - Transmission Risk Index;

  - Mortality Risk Index;

  - Testing Needs Index;

  - Medical Supply Needs Index;

  - Vaccine Needs Index;

  - Health Equity;

  - Variant Analysis / Prediction; and

  - Secondary Bacterial Index.

- Include our Climate Change and Zoonotic Spillover impact on Human Health model.

    - Human health risks and outcomes are impacted by:

        - Expansion/change of animal habitat due to flooding (and other natural disasters), hurricanes, and climate change [data can be sourced in part from EO data]; and

        - Changes in human habitat (land use change, population growth, etc.) [data can be sourced in part from EO data].

- Identification of areas of risk of flooding-induced vector-borne disease [data can be sourced in part from EO data].

- Creation of a Health-Disaster Vulnerability Index that conveys the situation on the ground:

    - Include information from environmental conditions.

- Conduct additional satellite imagery analysis in regard to health and health outcomes.

- Build out a UI/UX for end users, potentially in collaboration with:

    - StormCenter (GeoSolutions); and

    - Planet Crust.

- Identify replicable health-focused recipes for converting source data to ARD to DRI to indicators:

    - Inform partners of health data sources and recipes from source to indicators.

- Mobile Application:

    - Exporting of our data and health risk indices to existing mobile applications; and

    - Development of mobile application in conjunction with current and potential future partners.

- Access to First Responder input:

    - Coordinate focus groups and other mechanisms to gain insights from the first responder community related to information needs during a co-incident natural disaster and pandemic scenario

- Numerous additional potential collaborations have arisen that provide great value to the overall disaster response capability that should be considered in future pilots, including:

  - Planet Crust

    - build out of a UI/UX for end users; and

    - support efforts to make data input and risk index/indicator output (ARD, DRI, Indicator, Risk Indices) JSON-LD compatible.

  - Esri

    - Make Health/Pandemic Risk Indices available on ArcGIS.

  - GS1

    - Coordination of the tracking of the resources necessary in an emergency scenario.

  - GISMO

    - Develop survey language for the curation and collection of health-focused and broader ARD data relevant to a disaster scenario.

  - RSS-Hydro

    - Support feature extraction from EO data as relevant for health response needs.

  - Safe Software

    - Support data and risk index/indicator communications among end users.

  - SatCen

    - Support feature extraction from EO data as relevant for health response needs.

  - StormCenter

    - Communicate health risk indices for broad distribution.

  - Secure Dimensions

    - Further efforts to develop security and privacy workflows leveraging secure GeoPackage.

5.4. Analysis Ready Data Component (RSS-Hydro)

As part of the Disaster Pilot 2021 (DP21), the D101-4 Component Deliverable should provide service elements to support loading, preparation, and access to individual satellite-based observational datasets as well as other (e.g. demographic) datasets in forms suitable for analysis, visualization, and integration into decision-ready information products. The following data streams will be provided for possible flood hazard scenarios, to be picked up by services of this pilot. These data products are globally available, are in the form of ARD. Most are already in an OGC interoperable format. They comprise the following:

- Dartmouth Flood Observatory (DFO: https://floodobservatory.colorado.edu/) flood layers of events and historical maxima based on images from MODIS, S-1, Landsat and sometimes even commercial imagery;

- NOAA VIIRS operational flood map layer (daily availability);

- DFO satellite radiometry based river discharge estimates for a selection of "virtual" river measurement stations (daily availability);

- Free available global flood hazard model layers in Geotiff format from the EC Joint Research Center (JRC). These model layers are for different probabilistic flood return periods and show flood depths and extent at a pixel resolution of 90 m but are downscaled if possible for the area of interests (AOIs) of the pilot scenarios;

- Free global probabilistic rainfall forecasts from ECMWF NWP (https://www.ecmwf.int/); and

- GPM iMERG 3-hour rainfall estimates (https://gpm.nasa.gov/data/directory).

Figure xxx below illustrates a number of samples of such available datasets.

For those data layers not already available as such, the team can prepare any of these data in ARD formats and in OGC interoperability formats

Data (or part of the data) will also be made available, if desired, on GeoPlatform (https://www.geoplatform.gov/), which the team has successfully demonstrated during the 2019 Disasters Resilience Pilot (see demonstration video at: https://www.youtube.com/watch?v=tQJQT5xAGjc).

Any of these data can of course be provided for both the Canada and Peru flood hazard scenarios, or another use case location

Note that some of these layers have been showcased in the 2019 Disasters Pilot by the same team.

As part of the Pilot flood hazard scenarios and its D114-5 Component Deliverable (Mobile Analysis and Visualization Components), a free mobile client application (DFO-Floods) exists on the Apple and Google app stores and can be used to discover, request, and download decision

ready information products as GeoPackage in support of disaster response field personnel, operations, and decision making during connected-disconnected operations (Figure xxxx below).
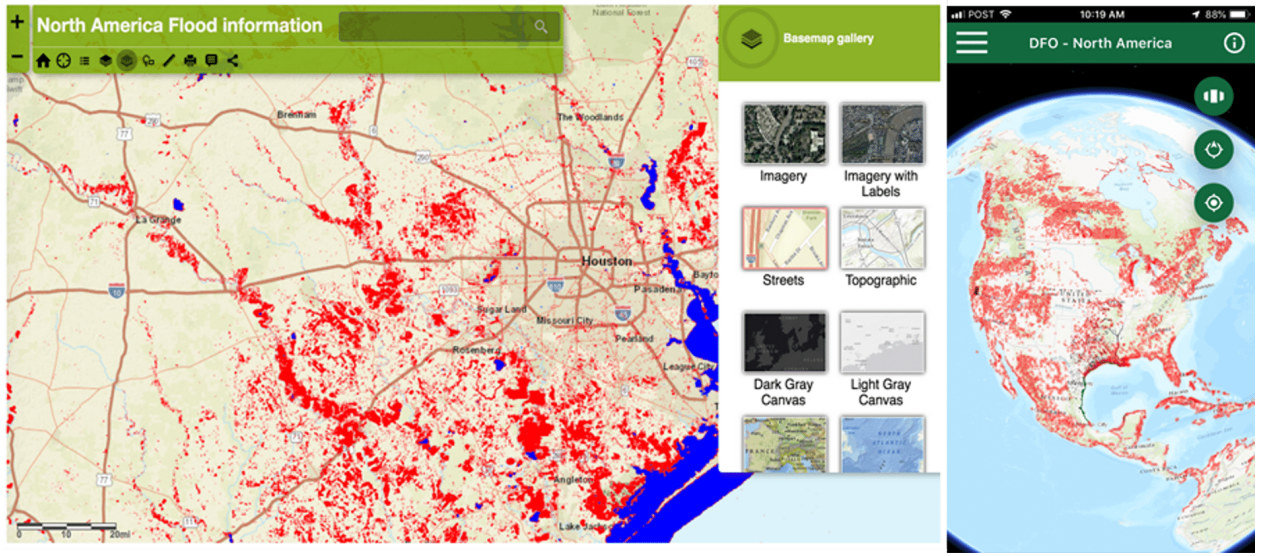
Figure 24 — This is an example figure

5.4.1. Relevant data sources

5.4.1.1. Red River of the North scenario

For flood impacts we took a closer look at some possible flood indicators. In particular, we looked at how different flood levels and extents would impact transportation infrastructure, and how this in turn could be used to generate routes around the flooding hazard. These recipes are reusable — they can be applied in different locations or for different times. Also, this recipe approach forms a pattern that can be applied to support other indicators related to population impacts such as evacuation zones.

This particular recipe is a collaboration between RSS Hydro, Dartmouth Flood Observatory, Safe Software, and Skymantics. RSS produces the flood time series based on DEM, flood gauge, and flood extents. DFO provides flood extents from sat imagery (MODIS). Safe Software provides a transformation and integration component based on FME which vectorizes and simplifies the raster time series. Skymantics provides the routing application which takes the IRD from Safe Software and uses it to support road condition indicators and routing requests.

This recipe was initially applied to the Red River scenario. The Red River of the North has a large basin in central North America that spans the US/Canadian border. The river flows from the Dakotas and Minnesota in the United States to Manitoba and Hudson's Bay in Canada, draining almost 300,000 sq kms. The Red River is susceptible to flooding due to its flat slope and typically large snow melts & ice jams in spring. One such flood was in the spring of 2011 was categorized as a 300 — 2000 year flood. The flood modeling was based on this historic flood.

5.4.1.2. Flood mapping with satellites

For Earth Observation monitoring of the flooding in the Red River basin, we used MODIS optical imagery with a ~250m resolution that provides two images a day from anywhere on Earth. MODIS data is excellent in mapping large scale flooding, and therefore less useful, for example, in monitoring urban flooding. The Figure below shows a set of sample data over the Red River of the North Area of Interest (AOI). Red areas on the image on the right show where standing water is detected by MODIS for the flood event in 2011.
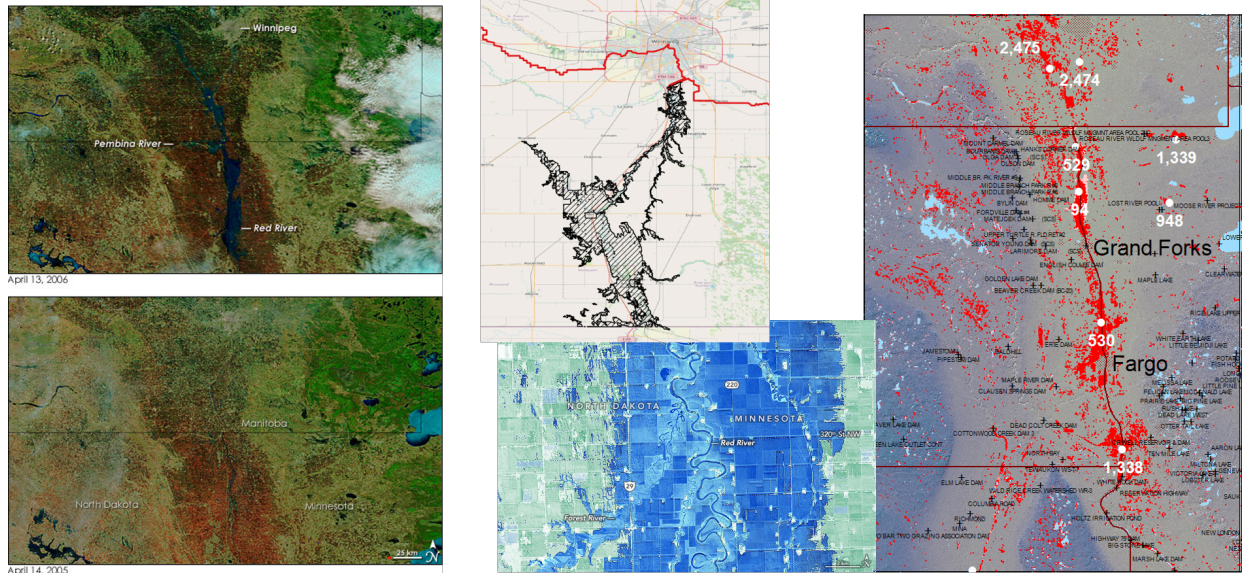
Figure 25 — This is an example figure

5.4.1.3. Flood modeling

The model time series was created using a high accuracy LiDAR DEM and a time series of measured water levels during the 2011 flood which were used to drive a simple "bath-tub" model that simulated close to 100 daily time steps of flood depths and extent dynamics (Figure Figure 26 below). These are delivered as GeoTIFFs to Safe's FME engine.
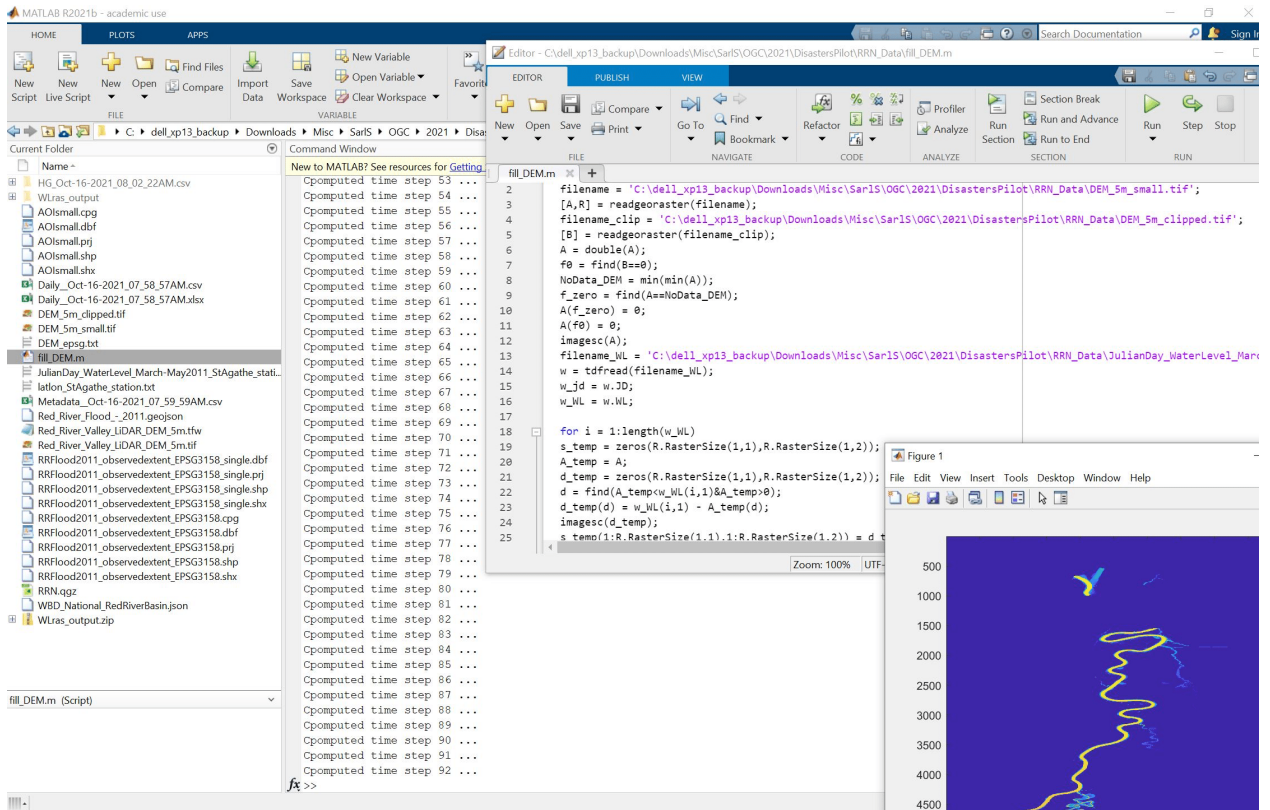
Figure 26 — This is an example figure

Once the flood time series was generated, Safe Software's FME component converted the data into Integration Ready Data to support Skymantic's routing engine and DRI. FME is Safe Software's data integration and automation platform that supports 500+ formats & APIs for data exchange including many OGC standards, along with a rich set of data transformation tools. It's also a no-code model based ETL approach that is ideal for rapid prototyping typically required in the disaster response context, and supports easy re-use. This particular transform reads the time series raster grid, classifies them according to the depth values needed, vectorizes & simplifies polygons, and then enriches the **polygons** with attributes. The result is a set of flood depth contours written to Geopackage and GeoJSON for use by downstream applications and published to the DP21 GeoNode.

5.4.2. Peru scenario

5.4.2.1. Extension of flood scenario to include climate change component

For looking at climate change effects, the frequency in observations about flood generating processes such as rainfall (known as return periods) can be correlated to the observed or simulated flood extent for the current conditions. Given the 21 IPCC climate model projections of rainfall, the changes in simulated flood extent by 2050 or 2100 can be determined and used to proportionally change the flood parameters such as extent and depth.

5.5. Analysis Ready Data Component (Safe Software)

Authors: Dean Hintz, Kailin Opaleychuk, Jenna Kim

5.5.1. Introduction

5.5.1.1. Pilot Overview

The OGC Disaster Pilot 2021 was an opportunity to test end-to-end information flow related to all phases of disaster management, with an emphasis on first responders, disaster coordinators, and other end users. Key to this was rapid sharing of interoperable data requiring minimum preparation in order to provide disaster response teams with geospatial information that makes a real difference to response activities. It is also crucial to define the data value chain for each indicator, from source data to Analysis Ready Data (ARD), Integration Ready Data (IRD), and Decision Ready Information (DRI), and make sure interoperability is supported across all the pilot components that contribute to this value chain.

5.5.2. Goals of participation

Supporting open standards is a very important part of our vision at Safe Software. This is because open standards are one of the key ways in which the integration and harmonization of data across disparate systems can be made easier. Ultimately this helps support enterprise-wide data sharing and automation, and breaks down barriers between data silos — which aligns with one of the central purposes of the FME platform.

Safe Software's involvement in the pilot focused on 3 principle application scenarios: Flooding in the Red River Valley, Pandemic in Louisiana, and Landslides in Peru. For each of these scenarios, Safe Software provided components related to the entire value chain as well as interacted with a number of components provided by the other participants. Safe Software's components and services supported data acquisition and transformation of this source data into ARD and IRD in order to support DRI results and also provided a basic web application as an example of how DRI results can be communicated to end users such as first responders on the ground. This section will detail both the inner workings of each of Safe Software's components, as well as how each component interfaced with others in the data value chain described above.

5.5.3. Technical design and implementation

Analysis Ready Data / Integration Ready Data Datasets Produced

5.5.3.1. 2011 Red River Flood in Manitoba, Canada

This scenario focused on flooding and flood impacts in the Red River Basin, an area prone to frequent flooding due to typically rapid spring melting, relatively flat slope, northward flow, frozen ground, and frequent ice jams. The Red River flows north from Northeast South Dakota and West Central Minnesota in the United States into Manitoba, Canada and eventually out into Hudson Bay. A typical spring thaw occurs from the middle of March across southern portions of the basin and mid or late April across the north.

After looking at a range of possible indicators, it was decided to develop indicators for best transportation routes to avoid flood water, along with recipes for the supporting data value chain. The value chain starts with observations of flooding using both river gauge measurements and satellite EO observations. These datasets are combined with mathematical algorithms and a Digital Elevation Model (DEM) which describes the height of the land, to produce two different approaches for the creation of a flooded area ARD.

5.5.3.2. Red River Pilot Recipe

Recipes describe workflows that transform ARD into IRD and AO to drive DRI's for specific critical management decisions. Recipes are designed to be reusable in different contexts: location (MB, LA, Peru) or time (past, present, future). The goal for the recipe developed for the Red River Scenario was to generate an indicator for best transportation routes to avoid flood water.

The main components for this recipe are as follows.

- Source data:

    - Hi resolution DEM for Red River Basin

    - Basemap data from OSM

- RSS Hydro & DFO: Flood Model & Extents

    - DEM + flood gauge + Modis extents ⇒ flood time series grids

- Safe Software: Grid to Vector Transform FME Model

    - Flood time series grids ⇒ Flood contours vector Geopackage

- Skymantics: Routing Web App

    - Flood Contours Vector Geopackage + Road Network + Tolerance rules ⇒ Flood mitigated routing

5.5.4. Scenario workflows and recipes

5.5.4.1. Source Data Preparation, ARD preprocessing

5.5.4.2. OSM to Geopackage

Before ARD and DRI can be generated, any pilot depends on some foundational base map datasets. These datasets also enhanced the ARD with more thematic and visual context and background for the decision ready indicators for our areas of interest. FME was used in the process of enhancing ARD for flood related disasters by reading in OpenStreetMap PBF data to generate Geopackage and GeoJSON outputs.

Before translation, the datasets were filtered and simplified to preserve the infrastructure and community related information such as roads, hospital locations, relevant amenities, and residential buildings. Through spatial and geometric filtering and schema mapping, this process was able to bridge any database limitations.
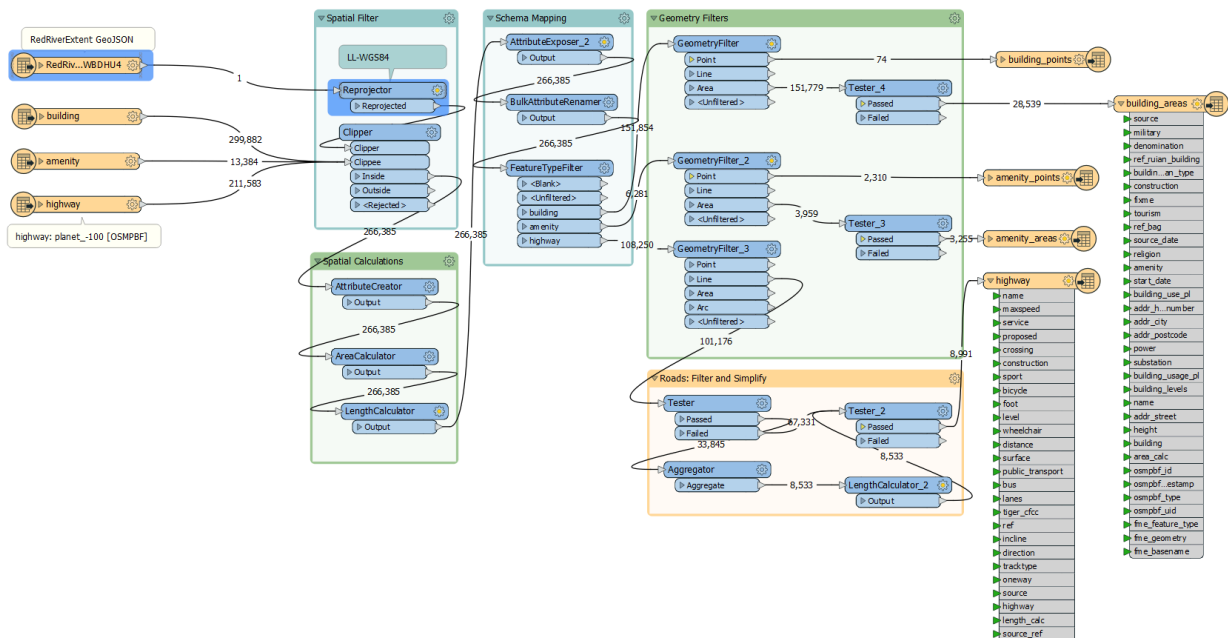
Figure 27

FME workspace for reading from OSM PBF and converting it to OGC Geopackage

Information was extracted from OpenStreetMap by manually selecting the area of interest. The extracted data was saved as a .OSMPBF file. FME's OSMPBF reader allows each of the layers to be read in separately for any data manipulation such as spatial filtering and schema mapping. Once the data was simplified and transformed it was written out as OGC compliant Geopackage for use by other components via file sharing services such as S3 or gitlab. The data was also uploaded to HSR.Health's GeoNode instance which allows this data to be available to other

components through OGC services such as WMS and WFS. The format of this data isn't limited to OGC Geopackage and can be delivered as GeoJSON and ESRI Shapefile as well.

Other ARD/DRI components of disaster preparation can benefit from utilizing information from these layers as disaster preparation needs to not only take account of the underlying geography but the human infrastructure built around and on top of it. This process of extracting data from OpenStreetMaps to shipping the information through appropriate formats is streamlined through FME's data manipulation capabilities.

5.5.5. Model based ARD

This first approach used the river gauge measurements and overlaid these onto a DEM and then a computer model was used to predict which areas would be expected to be flooded with those river measurements. The figure below shows the output grid for the area of the Red River Basin where the model technique developed by RSS Hydro has predicted the flooded area using data from 2011.

Area of 2011 flooding, colored light to dark according to flood day (start to end) overlaid on DEM (grayscale). Flooding model output from RSS Hydro and DFO.
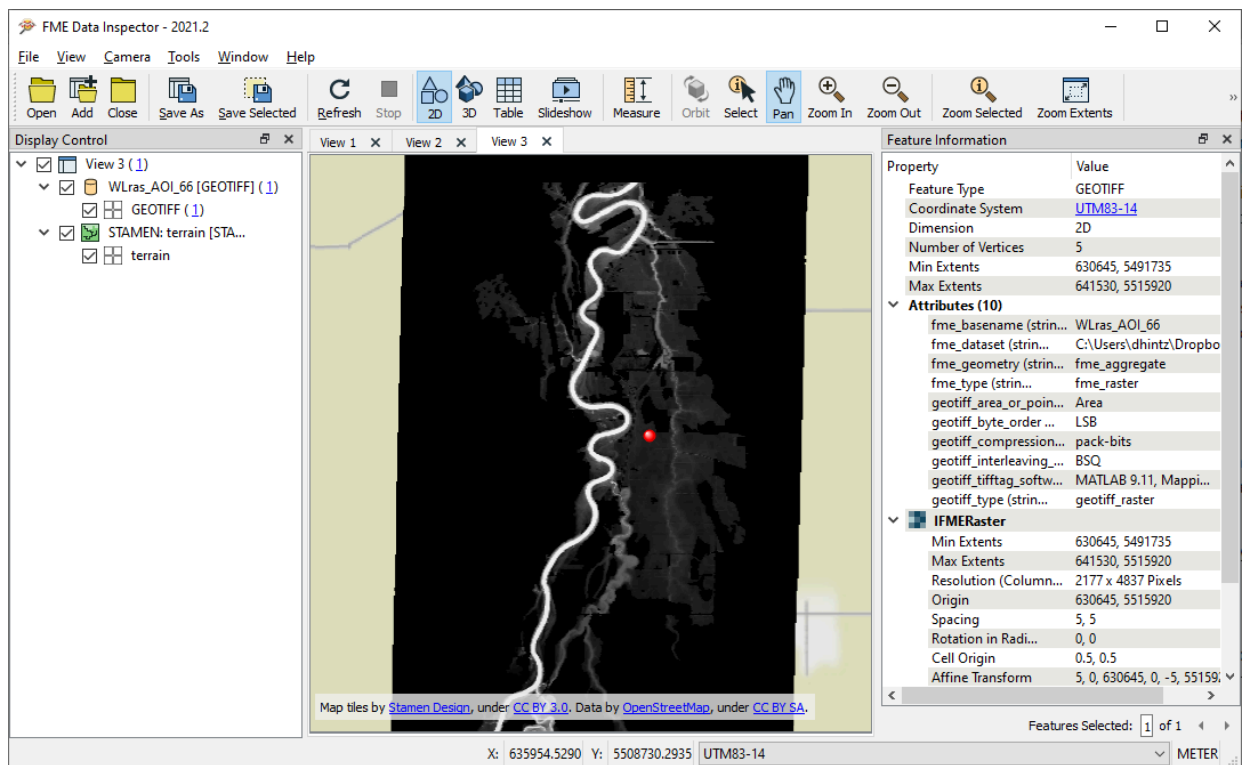
Figure 28

The above ARD grids of flooded areas were based on historic measurements. As such, they can be useful indicators of where flooding has occurred in the past and what the disaster teams can expect if similar water levels or rainfall are experienced again in the future, enabling them to get ahead of any flood response.

5.5.6. Flood Contours IRD with FME

As well as being used directly, the flooded area ARDs were useful along the data value chain and were used as input data for the next stage which transforms the data into Flood Contour IRD — integration ready data.

This approach was developed by Safe Software using the FME platform which is a model-based spatial data transformation and integration tool. Taking the input ARDs of modeled flood grids and EO ARDs, these were transformed into flood vector contour IRD, where areas with the same water depth were classified. This step was necessary as the transportation routing DRI, which was the next step in this data value chain, required flood depth estimates to work, not just the area or extents flooded. The flood water was categorized into five depth categories: 0.1 meters, 0.3 meters, 0.5 meters, 1.0 meter and above 2.0 meters.

Given the sensor and computational tools used, both EO and flood model output datasets tend to generate grid observations or time series. However, many analytic and GIS tools work more readily with vector datasets. This is why the ARD to IRD approach for flood impact analysis was designed to convert raster flood depth grids to vector flood contour polygons to better support downstream integration required for the IRD. The figure below shows an example of the workflow used for the FME part of the ARD to IRD / DRI recipe for this flood transportation indicator.
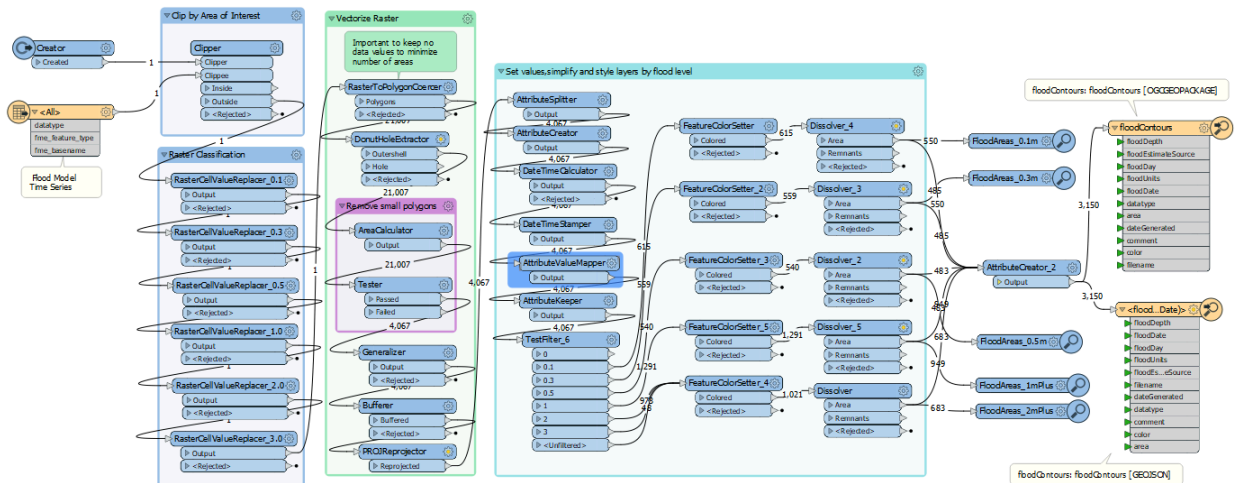
Figure 29

FME approach for converting flood area grid ARDs to flood contour IRD:

1. Raster classify, vectorize to polygons, simplify geometry, calibrate, reproject & enrich;

2. Generate IRD: integration ready outputs: Geopackage and GeoJSON;

3. Complete model run for input 2011 flood model 70 timesteps ⇒ output 200K records; and

4. Post to GeoNode (GeoServer).

To improve accessibility, the result was saved as an OGC GeoPackage, which makes it easy to share with other components as well as to use offline. The IRD Flood Contours were then used as input data for the creation of the transportation routing DRI. The figure below shows flood contours in the Red River Valley south of Winnipeg, which was used as the input data for the transportation route DRI.
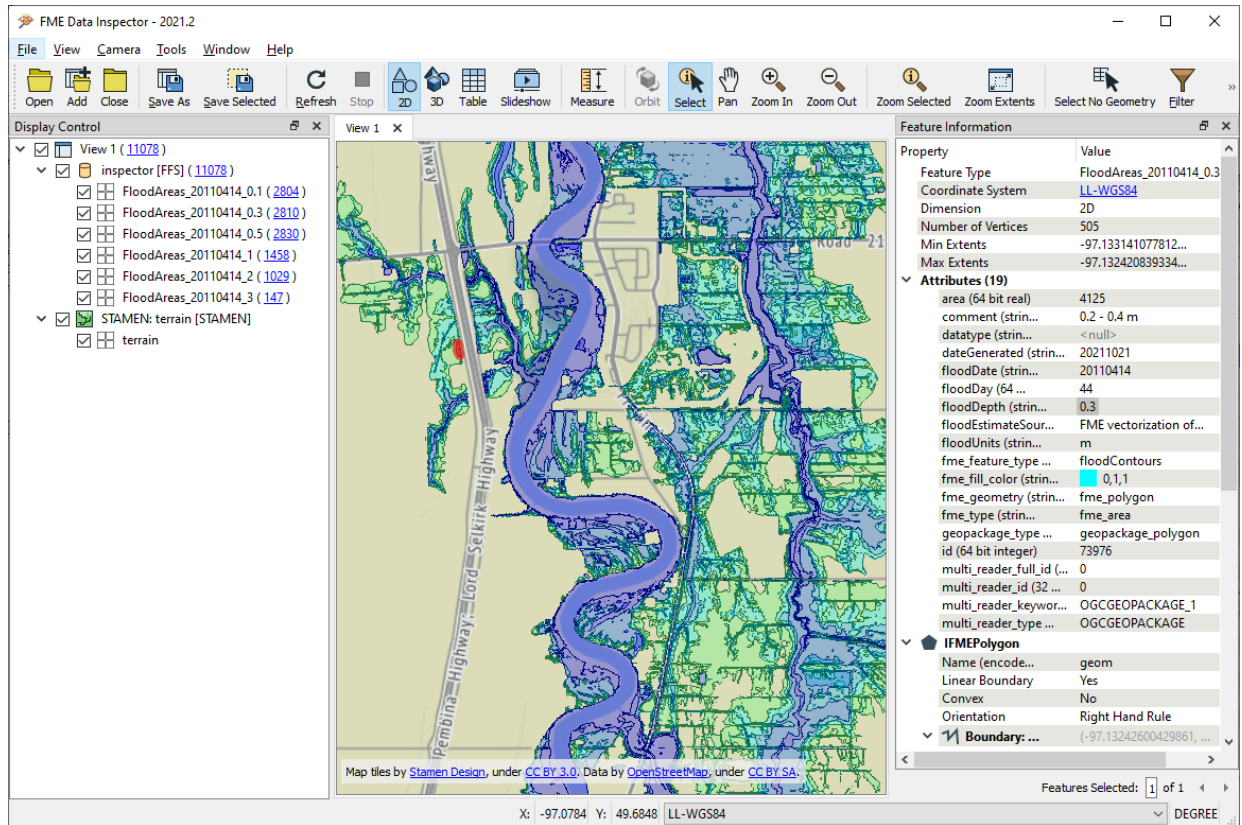
Figure 30

Flood Contour Geopackage showing flooded areas south of Winnipeg by date, as displayed in FME Data Inspector.

To better support online integration, the flood contours were also provided to the HSR.Health GeoNode instance, which makes this data available to other components via OGC services such as WMS and WFS. The figure below shows an example of the flood contours for the Red River Flood from 7th April 2011.
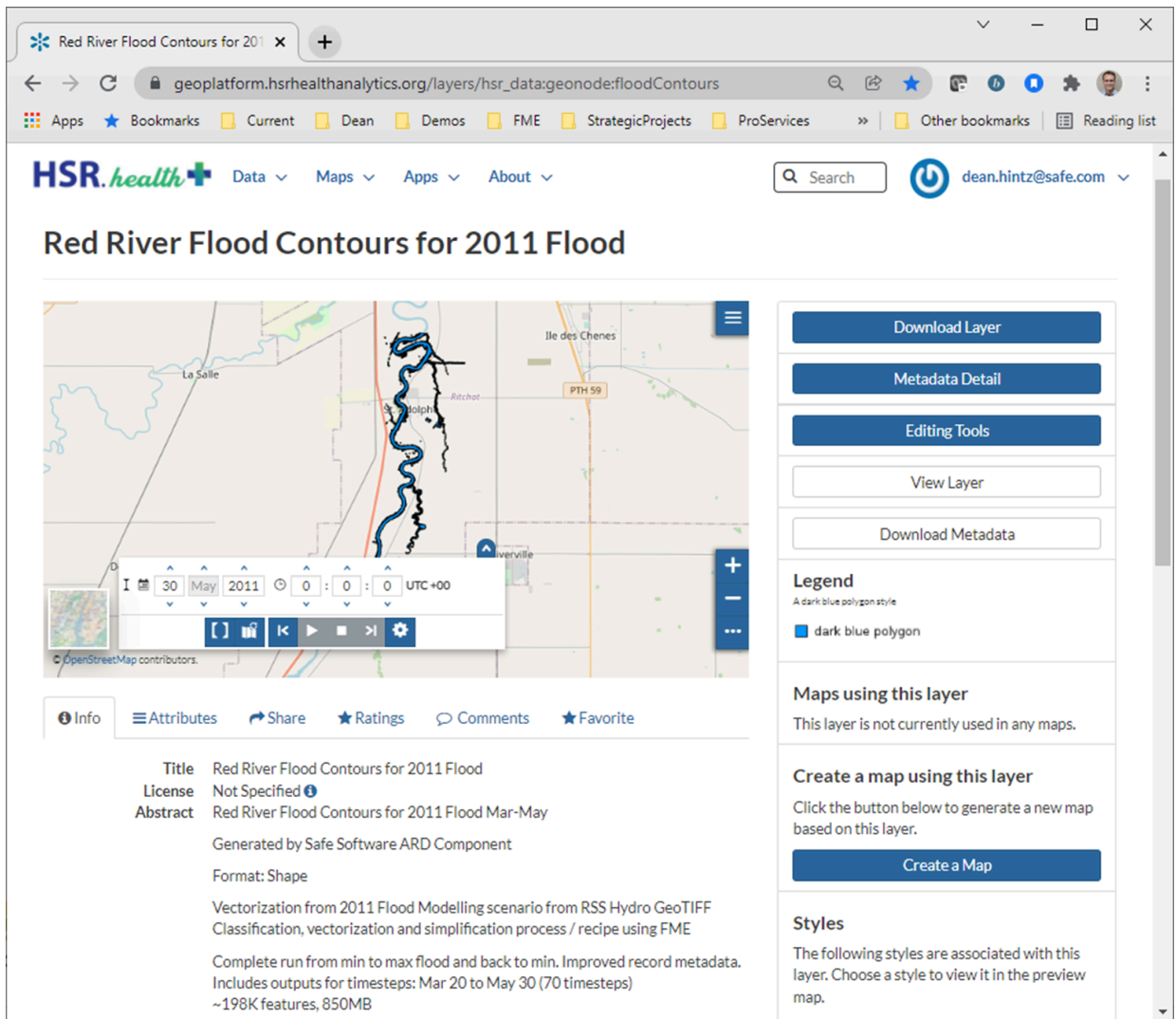
Figure 31

HSR Health Geonode with Flood Contours for Red River flood from 7th April 2011 loaded from FME workflow outputs.

Note that the DP21 Pilot GeoNode supports: layer search, metadata, styling, interactive web map interface, time series animation, data download, and OGC web services WMS and WFS.

5.6. Decision Ready Indicator Component (Safe Software)

5.6.1. DRI — Decision Ready Indicators for Flood Aware Transportation Routing

For the main transportation impact indicator developed for the Red River scenario, the DRI component was implemented by Skymantics and is briefly outline here in order to explain the overall process chain for this scenario. Other test DRI components were implemented by Safe Software and are described after the transportation impact component.

The ARD component above produced the vectorized flood contour ARD / IRD dataset needed to support downstream flood impact indicators. The Skymantics transportation DRI component then took this data and used it to calculate the transportation routing DRI. It used the flood contours depths to determine which roads were impacted by the flooding and how deep the water was for each road segment. The user had the potential to specify what depth of flooding is passable, which was important for different vehicles that could be involved. For example, a large lorry or 4X4 would likely be able to handle deeper water than a motorbike or small car. The user entered the starting position and the intended destination, and the routing software worked out the best route to take to avoid the flooded routes.

The diagram below demonstrates the process. Top left shows the optimal route between Ste. Agathe and Ile des Chênes, two places in Manitoba that are approximately 30 km apart. Top right shows the flood contours produced in the process above for Winnipeg and surrounding areas. Bottom left is the user specified elements, which in this case allows a maximum flood depth of 0.2 m for public vehicles and 0.4 m for emergency vehicles. Finally, in the bottom right is the new optimized route taking account the flooding and user requirements.
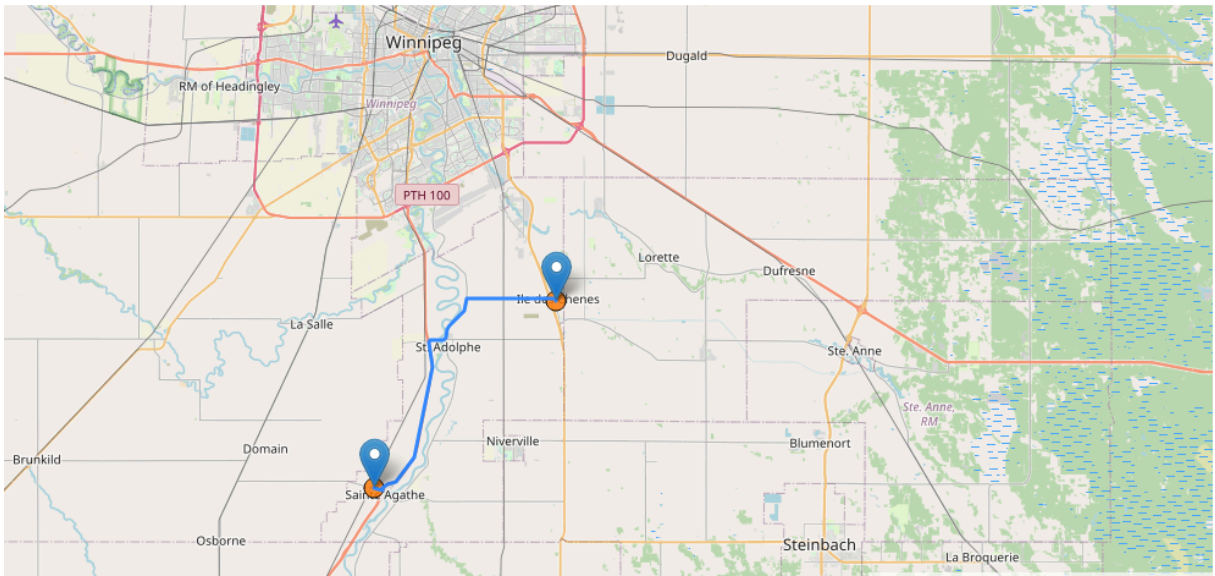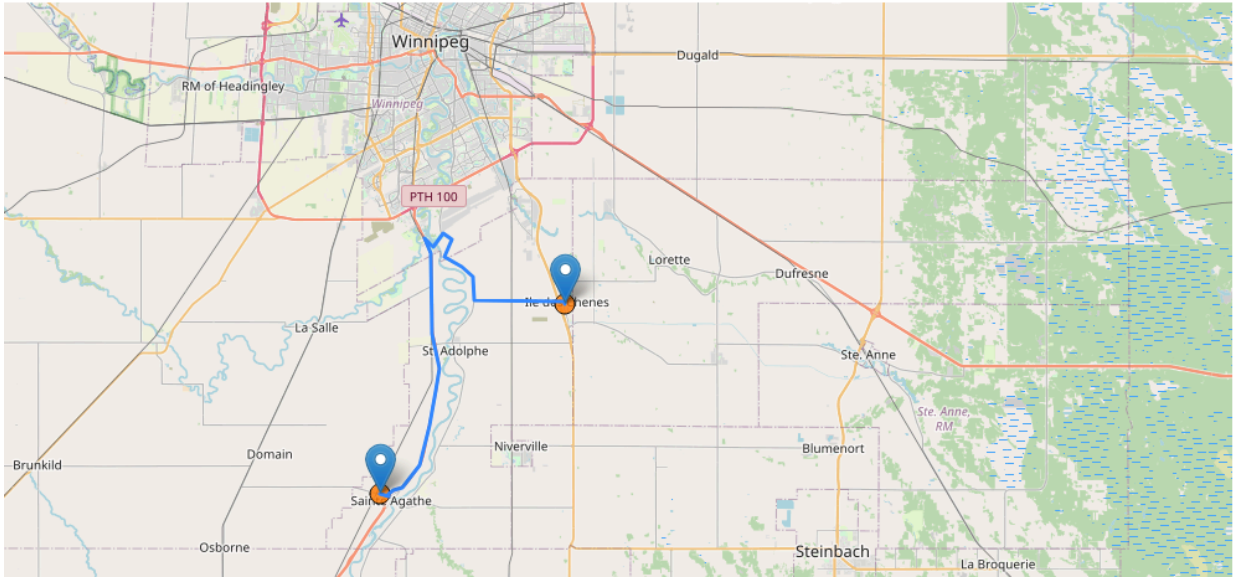
Figure 32

Figure 33

Figure 34

Skymantics Routing DRI showing optimal route, flood contours, user variables, and revised route (Produced by Semantics).

The routing software used the flood contours, topography of the area, and a road data set to find the roads which would be impacted by the flood, and then determined where the road was passable, passable only by emergency vehicles, or impassable. It used these determinations to create a revised route between the two locations. This type of application is important for response teams trying to get to situations, evacuation routes, or the delivery of supplies.

5.6.2. Red River Scenario Summary

This scenario has focused on flooding within the Red River Basin, Canada. The Pilot generated several ARD datasets using model predictions, satellite observations, and flood contours and then developed a series of DRI recipes. The Pilot then used the recipes created to produce a data value chain focused on the rerouting of emergency vehicles taking into account flood levels.

This scenario has shown that it possible to deliver the value chain to create the example DRI, however it has also identified some issues and limitations around automating the process.

Components from 4 participants successfully exchanged data via open standards to transform ARD into IRD and AO to drive disaster response DRI for transportation flood mitigation. Other recipes for flood impact indicators or other contexts could be added using this pattern.

This recipe is reusable in other contexts as described below in the Peru and BC scenario sections. Overall this scenario demonstrated that a standards based approach (OGC) was an important part of supporting interoperability & automation. GeoTIFF, GeoJSON, and

Geopackage allowed data to be readily exchanged between pilot components. Data was uploaded to GeoNode and published via OGC services.

For more information on the FME applications behind the Red River Scenario see: https://community.safe.com/s/article/OGC-Disaster-Pilot-2021.

5.6.3. Peru Scenario

We attempted to adapt the Red River transportation impact recipe to the Piura Basin in Peru. However, the main barrier to this was the limited or low resolution of the available DEMs for this region. This did not allow for the running of a similar flood model to the Red River scenario. It also meant that meaningful flood depth contours could not be generated, only flood extents as derived from SAR imagery.
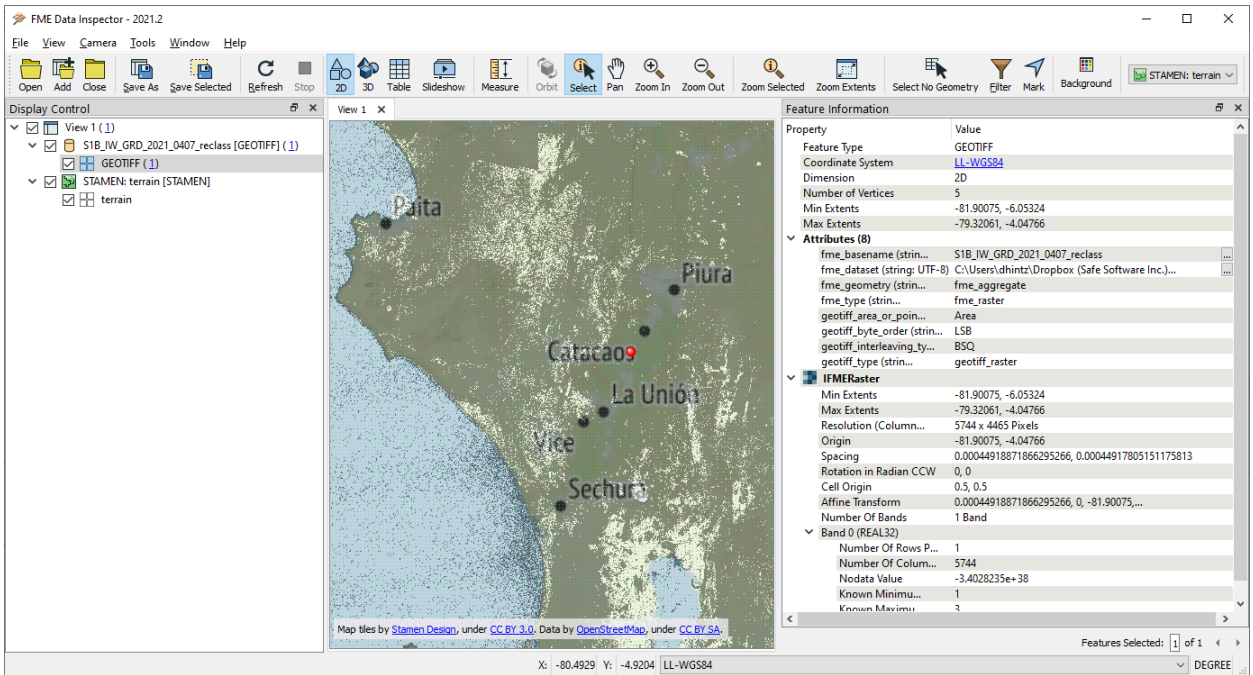
Figure 35

RSS Hydro: Sentinel-1 SAR of Flooding — April 7, 2021
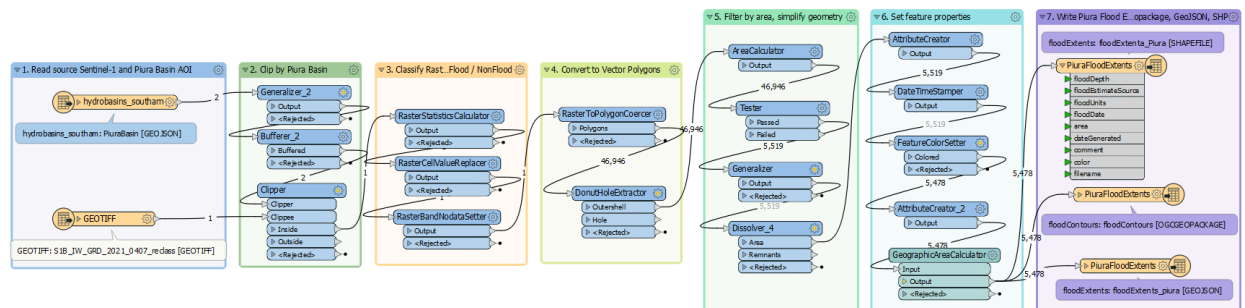
Figure 36

FME Recipe for processing Sentinel-1 data to generate vectorized flood extents for the Piura Basin:

1.   Read Source Sentinel-1 and Piura Basin extents;

2.   Clip by basin extent;

3.   Classify raster cells for probable flood, set no-data;

4.   Convert to vector polygons;

5.   Filter by area and simplify geometry;

6.   Set feature properties and metadata; and

7.   Write to destination formats and services.
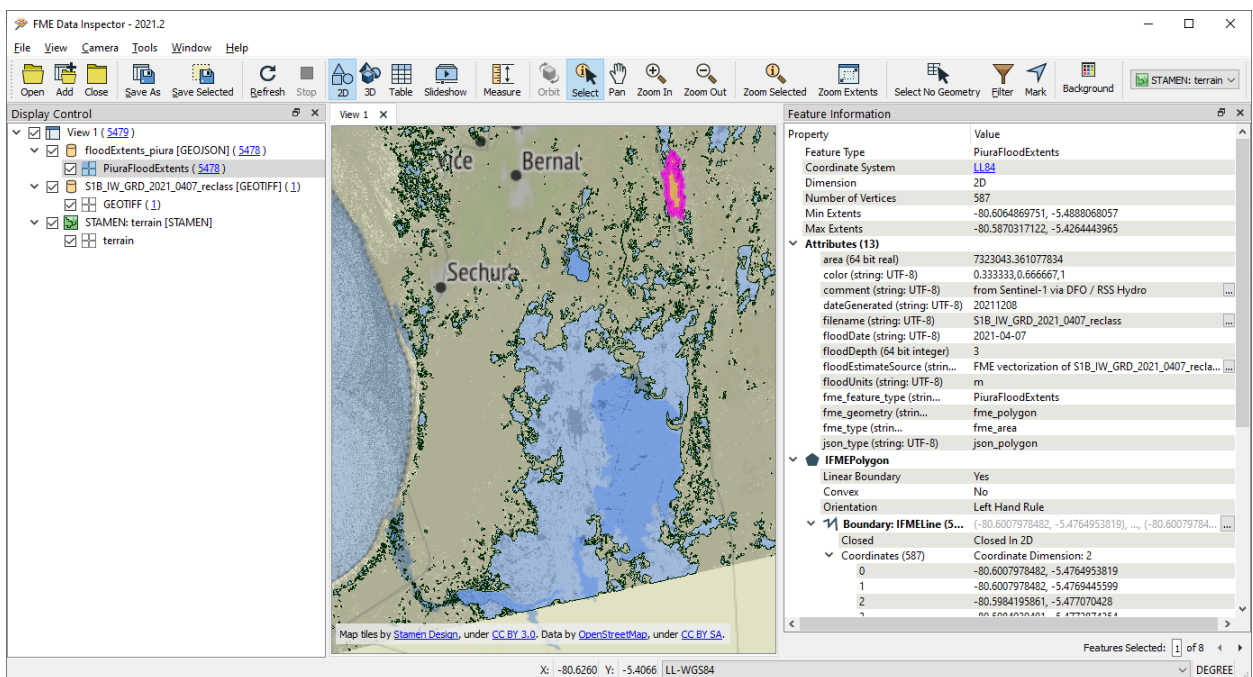
Figure 37

Piura Vector Flood Extents IRD: Geopackage, GeoJSON
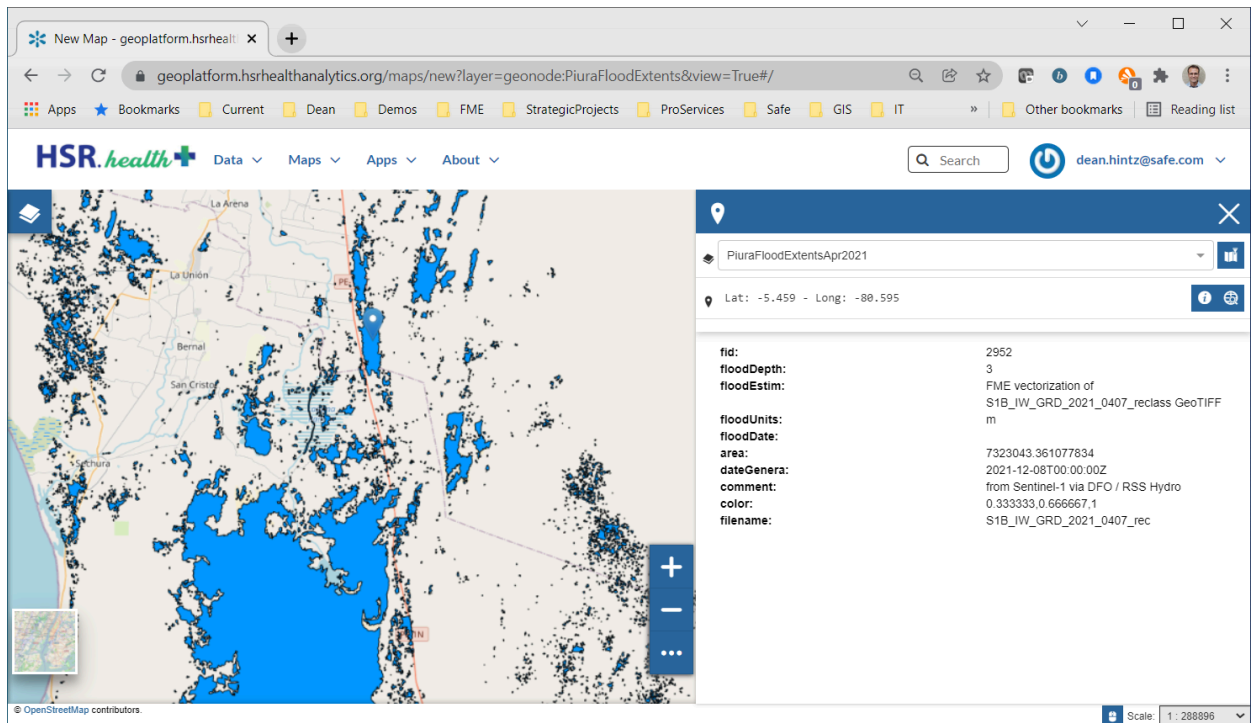
Figure 38

Piura Vector Flood Extents IRD: Published to GeoNode

5.6.4. British Columbia (BC) Flood and Landslide Scenarios

In the midst of the disaster pilot, BC experienced one of the worst natural disasters in its history with the extreme weather, floods, and landslides caused by Pacific storms in late November, 2021. This involved multiple cascading hazard impacts at multiple locations over time: landslides, pandemic, loss of infrastructure, and flooding. BC was also cut off from the rest of Canada; most major transportation routes were closed for more than a month. While the type of natural hazards are well known in BC, the scale and extent of local impacts were a complete surprise for many.

Figure 39

Sumas Prairie Flooding — CBC photo

One of the central goals of the pilot was to produce workflows and recipes that are readily adaptable to a wide range of contexts with minimal effort. Because of this, Safe Software tested how adaptable its recipes and workflows were by applying them to the BC flooding and landslides.

One key aspect for any rapid disaster response is mature spatial data infrastructure. Especially in the context of flood response and landslides, good baseline data such as base maps and high resolution DEMs are crucial. Fortunately for the BC context, Natural Resources Canada (NRCan) had generated a rich set of high resolution digital terrain maps based on a recent LIDAR survey of the area. These DEMs were publicly accessible from: https://www.nrcan.gc.ca/science-and-data/science-and-research/earth-sciences/geography/topographic-information/whats-new/high-resolution-digital-elevation-model-hrdem-generated-from-lidar-new-data-available/23005.
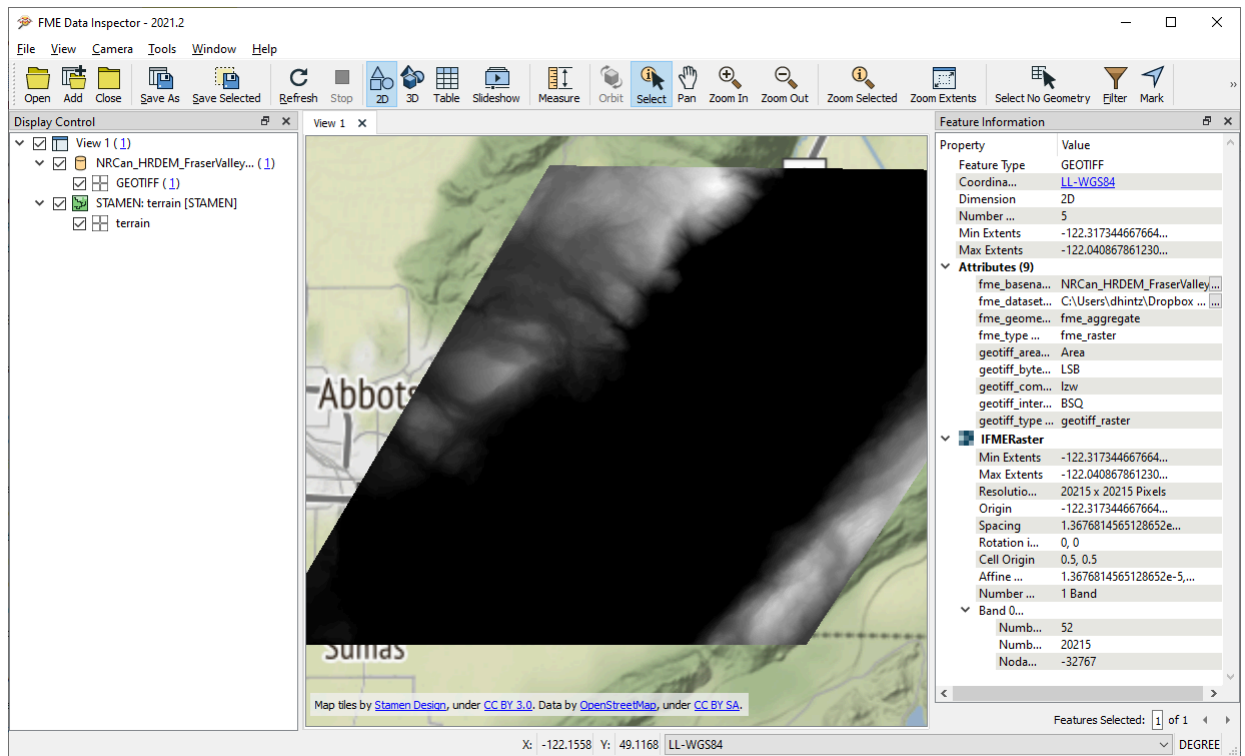
Figure 40

HRDEM NRCan. Reprojection and mosaic with FME

NRCan also provides satellite based monitoring for flooding across Canada based on imagery from RADARSAT. FME was used to download these datasets, convert them to GeoJSON, and geopackage and upload them to the HSR's GeoNode for use in downstream processing.

See: https://open.canada.ca/data/en/dataset/9cad712a-5ac5-4248-b7d7-2db1a3892509
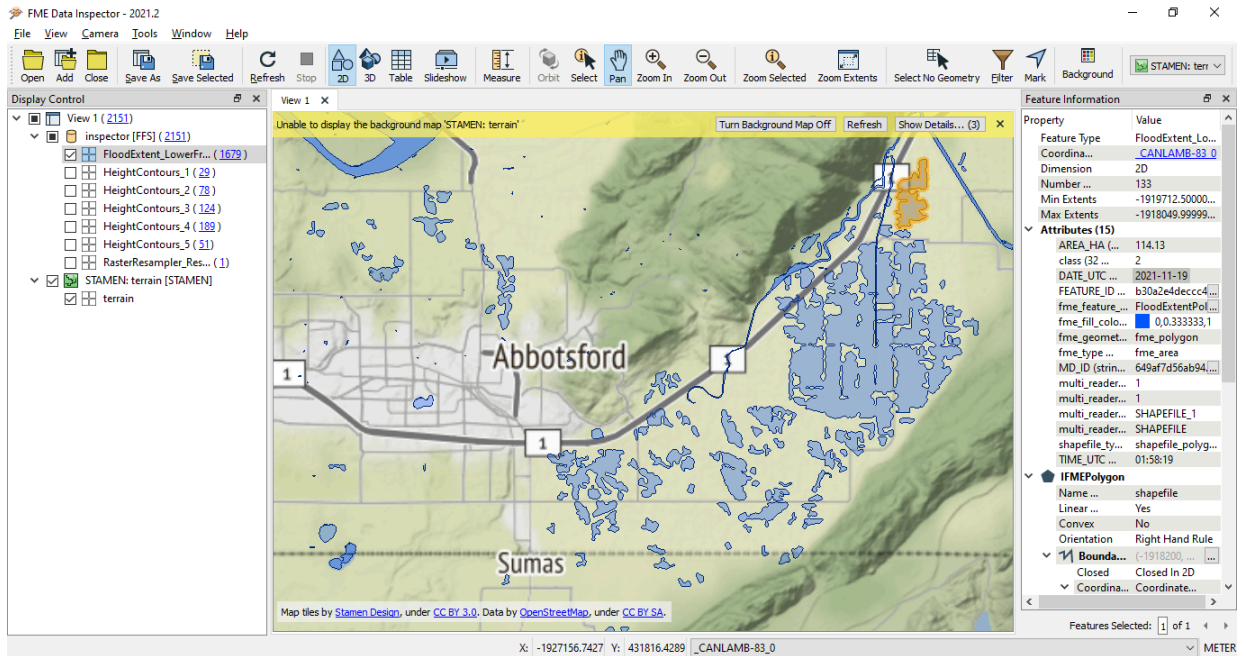
Figure 41

Fraser Valley Flooding — proximity to Abbotsford, and border at Sumas, WA

5.6.5. Flood & Landslide ARD — Raster Mosaicking

High resolution DTM tiles were obtained for the Lower Fraser Valley from Natural Resources Canada. FME was used to mosaic DTM tiles for the lower mainland of British Columbia to feed flood risk analysis and identify landslide hazards in the eastern Fraser Valley. The FME data preparation workflow involved reprojecting from CSRS to LL84, resampling, and mosaicking. Natural Resource Canada provides tiles in a 1m resolution, but FME's RasterResampler transformer resamples tiles to 2m to create a manageable output. If a raster extent or tile overlaps during the mosaicking process, any Nodata values are replaced by real data values from under or overlying raster. Finally, the mosaic is reprojected to LatLong84 before any clipping or slope calculations occur.
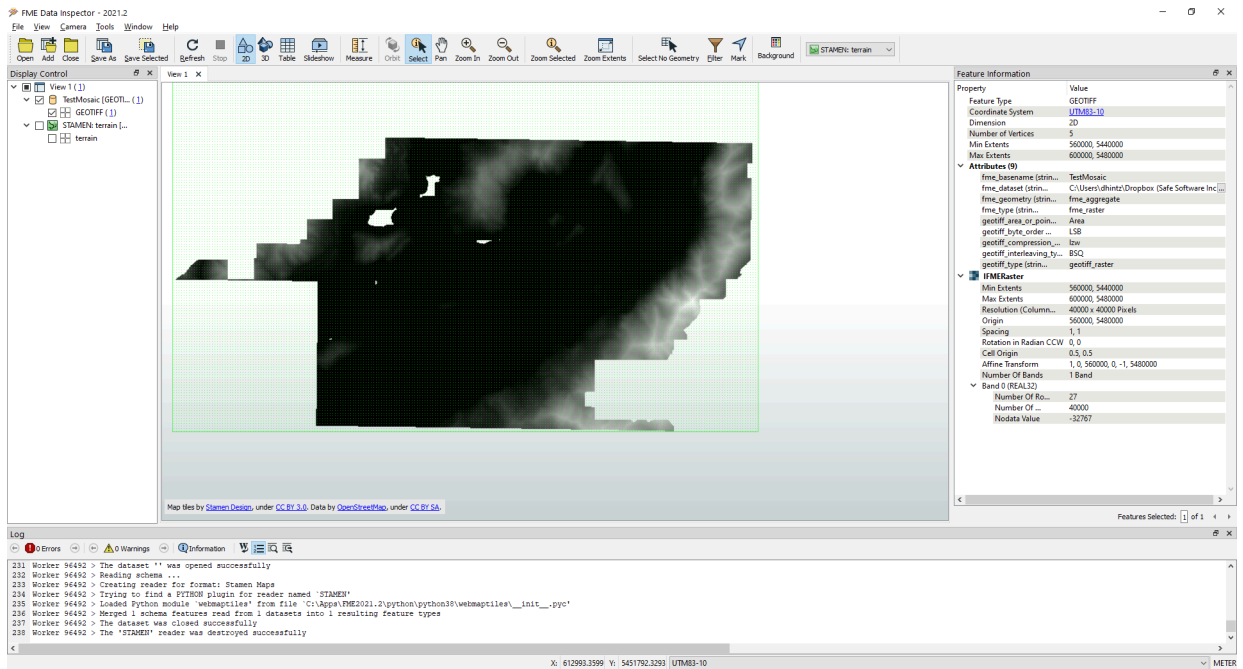
Figure 42

5.6.6. Fraser Valley Flooding — ARD IRD Process

A central pilot goal was to provide components that are able to rapidly rerun recipes in near real time. In order to better understand flood risks in the context of flooding in the eastern Fraser Valley, the Red River flood impact recipe was reused in BC context. The Red River vectorization FME process was rerun using HRDEM from NRCan and assessed alongside the current flooding in the lower Fraser Valley. When overlaid with NRCan active flood extents from RADARSAT a close correlation could be observed for local areas in the Sumas Prairie < 4m in absolute elevation. While not predictive, clearly these local low elevation areas, assessed from a high resolution DEM in the context of what was a historical lake, helps illustrate what areas pose significant risk for current and future flooding.
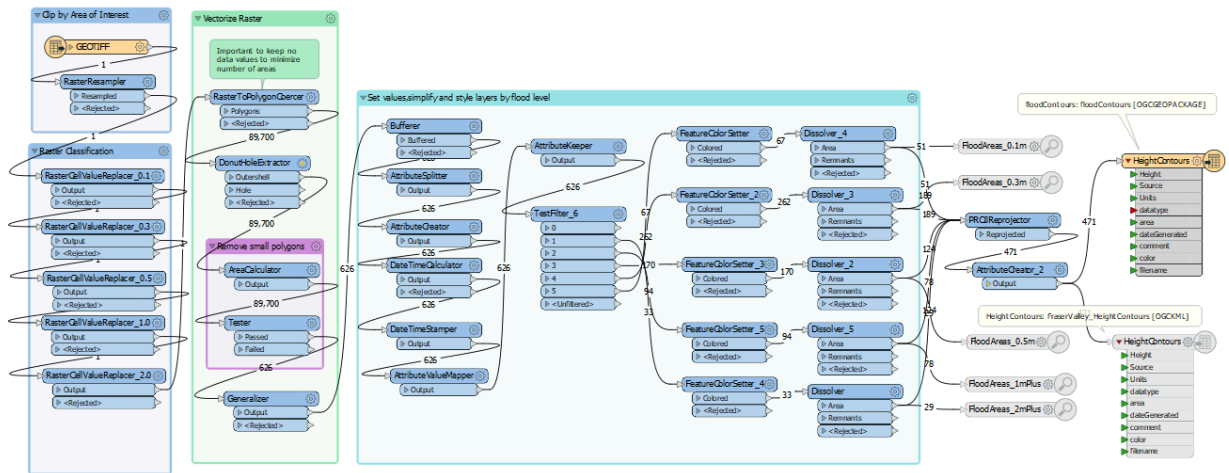
Figure 43

FME Conversion of DEM Grids to Flood Hazard Contours Vector Geopackage

This workspace demonstrates raster to vector conversions, followed by the removal of small polygons, generalization to reduce detail (e.g. by smoothing lines), classification into five depth categories (0.1, 0.3, 0.5, 1.0 and 2.0+ m), and finally dissolving (aggregate touching features to further reduce complexity) before writing to the Flood Hazard Contours geopackage.
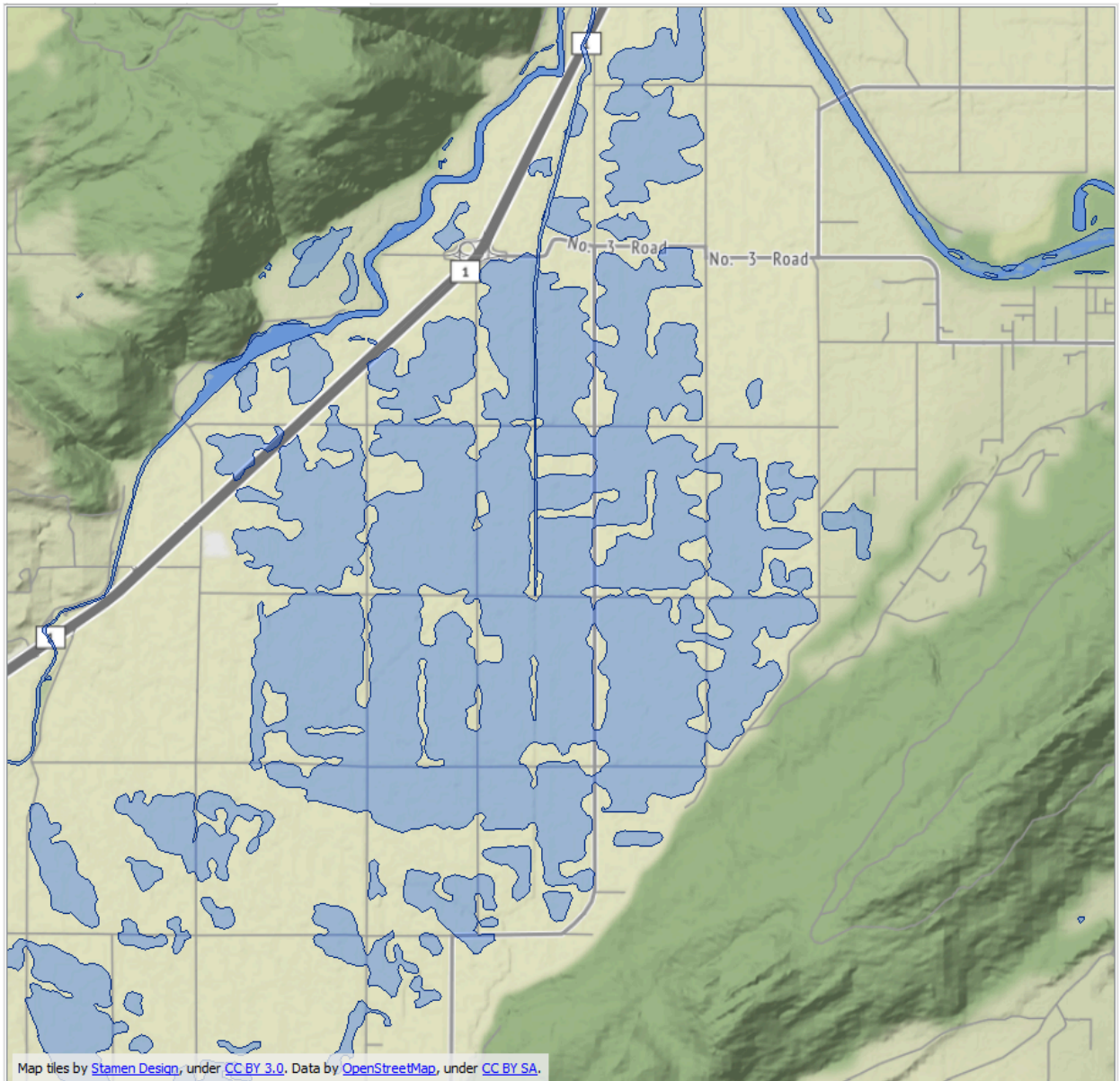
Figure 44

Nov 19 flood extents in Sumas Prairie, Fraser Valley, from Natural Resources Canada: active floods (RADARSAT)
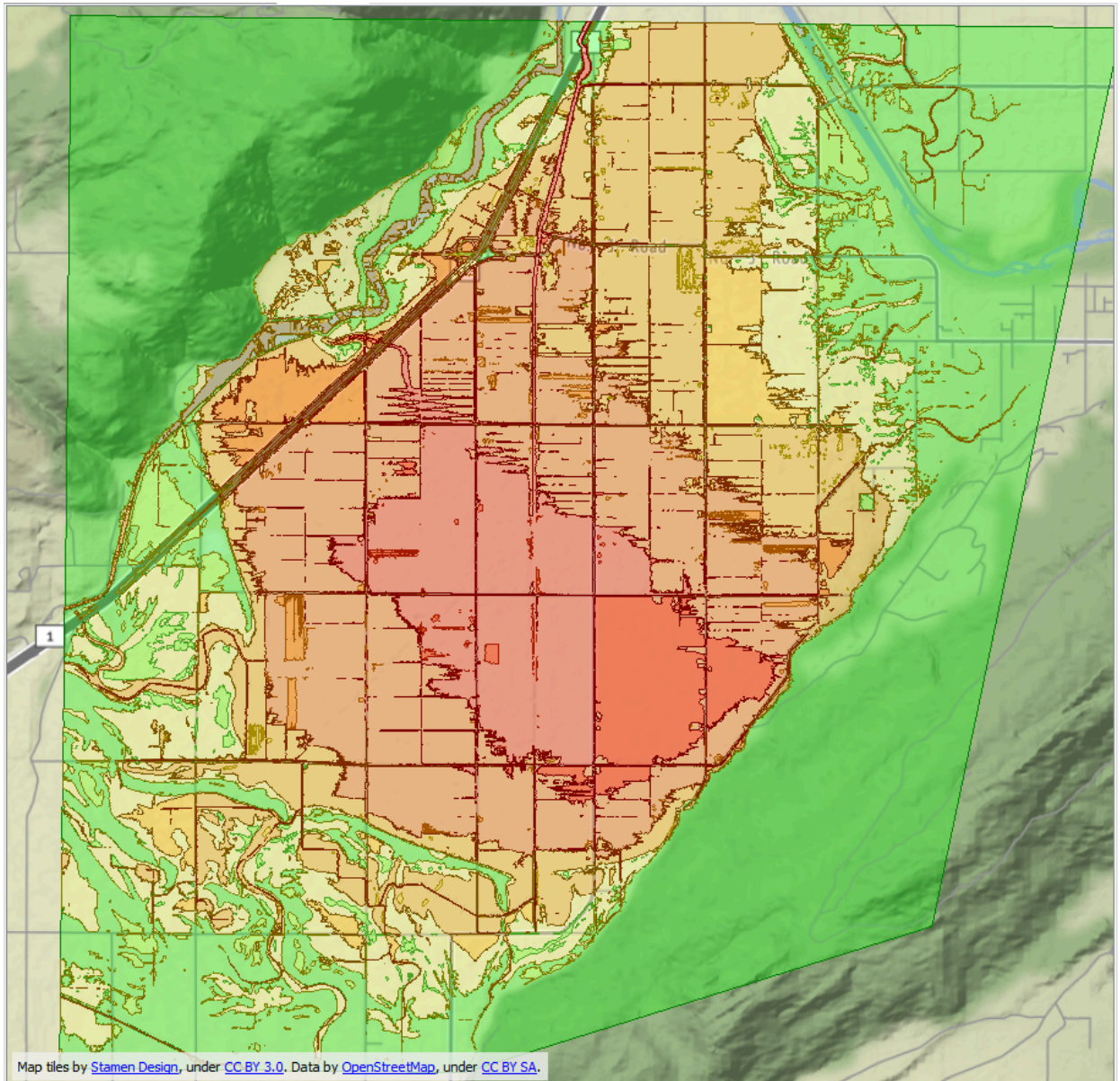
Figure 45

LIDAR DEM (NRCan) reprojected, classified, vectorized for Sumas Prairie, Fraser Valley
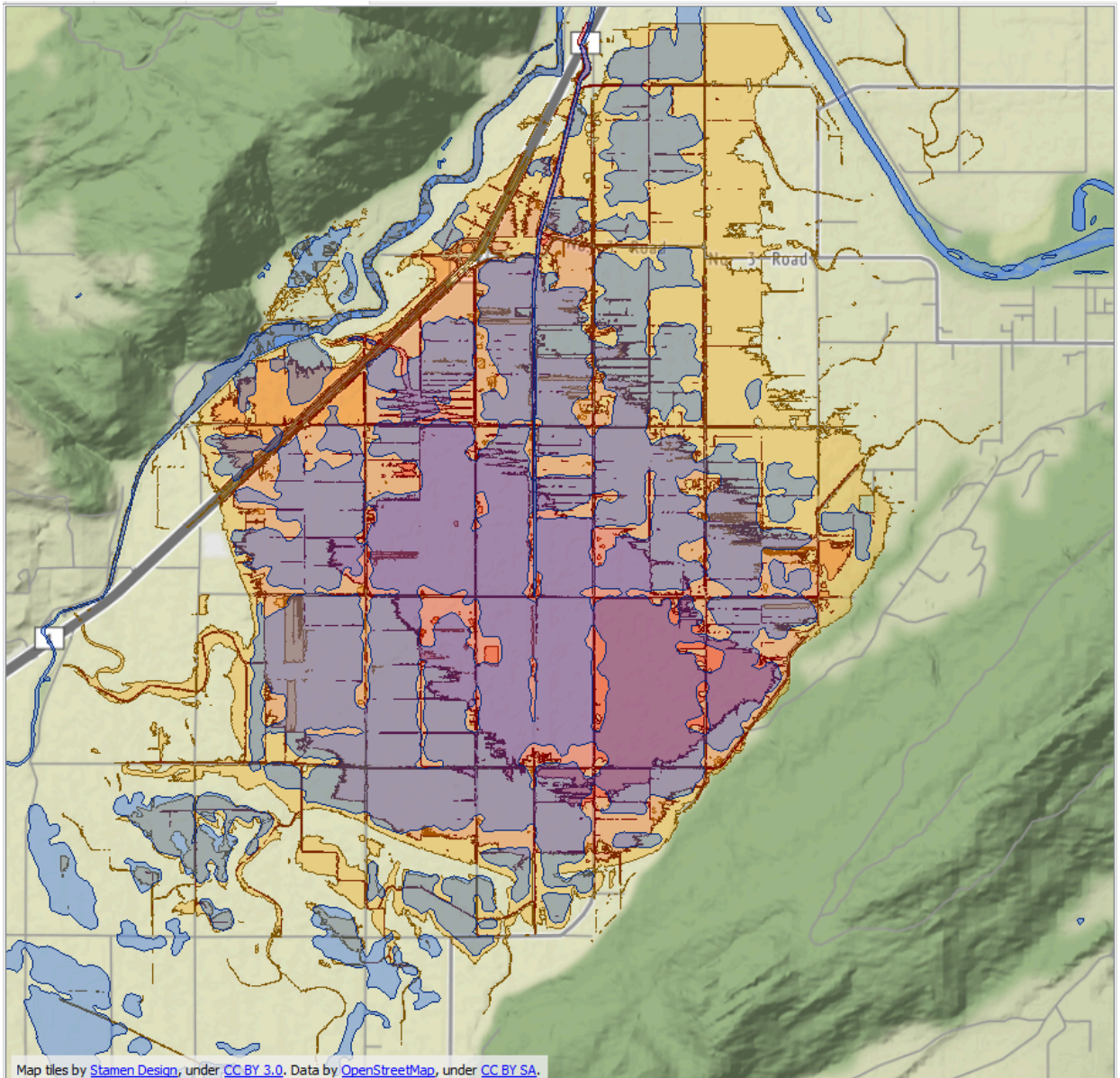
Figure 46

Overlap of < 4m elevation with Nov 19 BC flood extents (processed by FME component originally used for Red River, MB)

5.6.7. Landslide ARD

Given that landslides played a major role in the BC extreme weather disasters of Nov 2021, Soft Safewar looked at some basic modeling for landslide risk and compared some key factors with the actual landslide locations.

Some well-known key landslide risk factors are slope, rainfall, vegetation cover, geology, use, and changes to any of these such as recent land use or fires.

As mentioned above, one of the major challenges in the Peru context was the lack of availability for high resolution DEM. Another challenge unique to landslides is the highly localized nature of the hazard. For example, the precise location of tree cover, underlying geology, and ground seepage can have a large impact on whether or not a landslide occurs at any given location. Flooding, on the other hand, can cover a large area of equal elevation.

Figure 47

SAR crews assisting stranded motorists isolated by multiple mudslides along Highway 7 in the Easter Fraser Valley — Nov 2021 Image credit: Globe and Mail

5.6.8. Landslide ARD — Raster to Vector

As a very basic estimate of flood risk, the Hi resolution Fraser valley DEM was used to generate a grid of slope values. The slopes were then classified to look for steep slopes in close proximity to major highways. These areas then could be targeted for prioritized monitoring and protection.

A similar FME workflow to the Red River and Fraser Valley flood scenarios was used, but instead of classifying flood or elevation, the workflow to classify the slope grid into slope contours, or areas of common slope, was used. Areas with high slope (> 40 degrees) in close proximity to major highways like highway 7 or highway 1 were located.
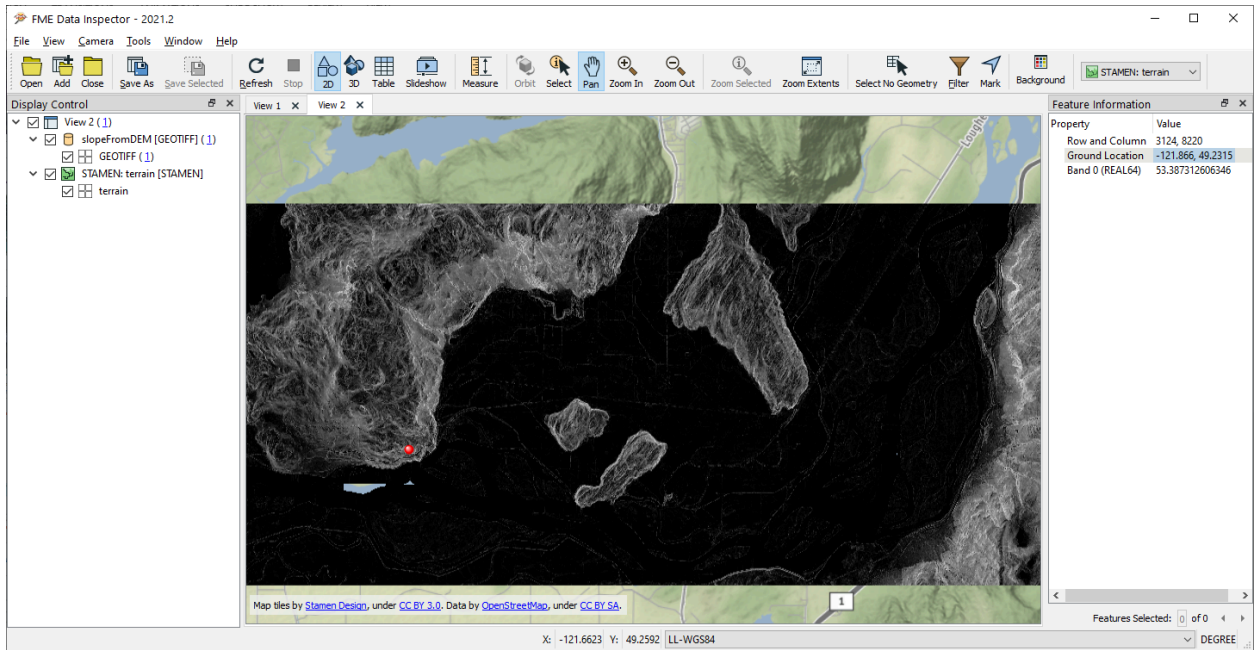
Figure 48

Slope Grid Derived from NRCan High Resolution DEM
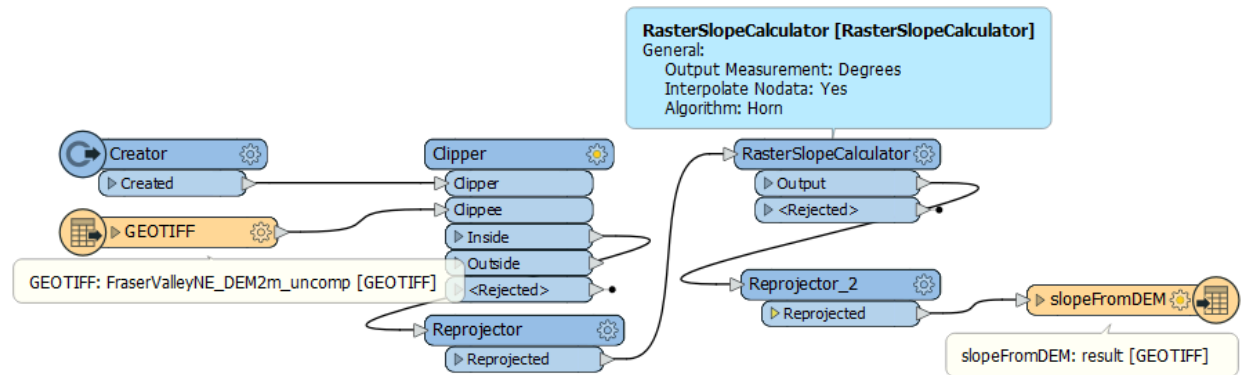
Figure 49

Slope calculation FME workspace
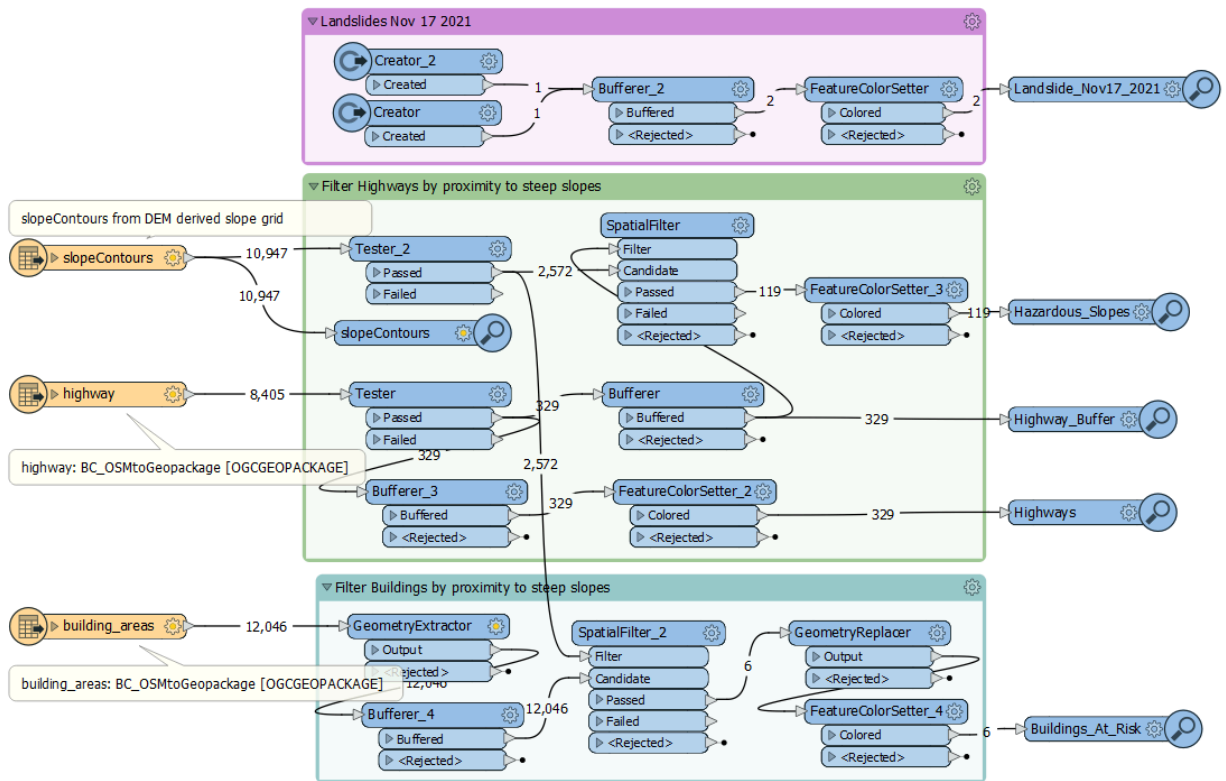
Figure 50

Landslide Risk Indicator Trial Recipe: Slope Contours to At Risk Highways

This workspace demonstrates raster to vector conversions followed by the removal of small polygons, generalization to reduce detail (e.g. by smoothing lines), classification into slope based categories (0, 10, 20, 30, 40, and 50° degrees), and finally dissolving (aggregate touching features to further reduce complexity).

Input: Slope Contours geopackage, Infrastructure geopackage with OSM roads & buildings

Process: Road type = highway or motorway, Slopes > 40 degrees, Proximity to highways < 120 m

Output: Highway areas within zones of higher landslide risk
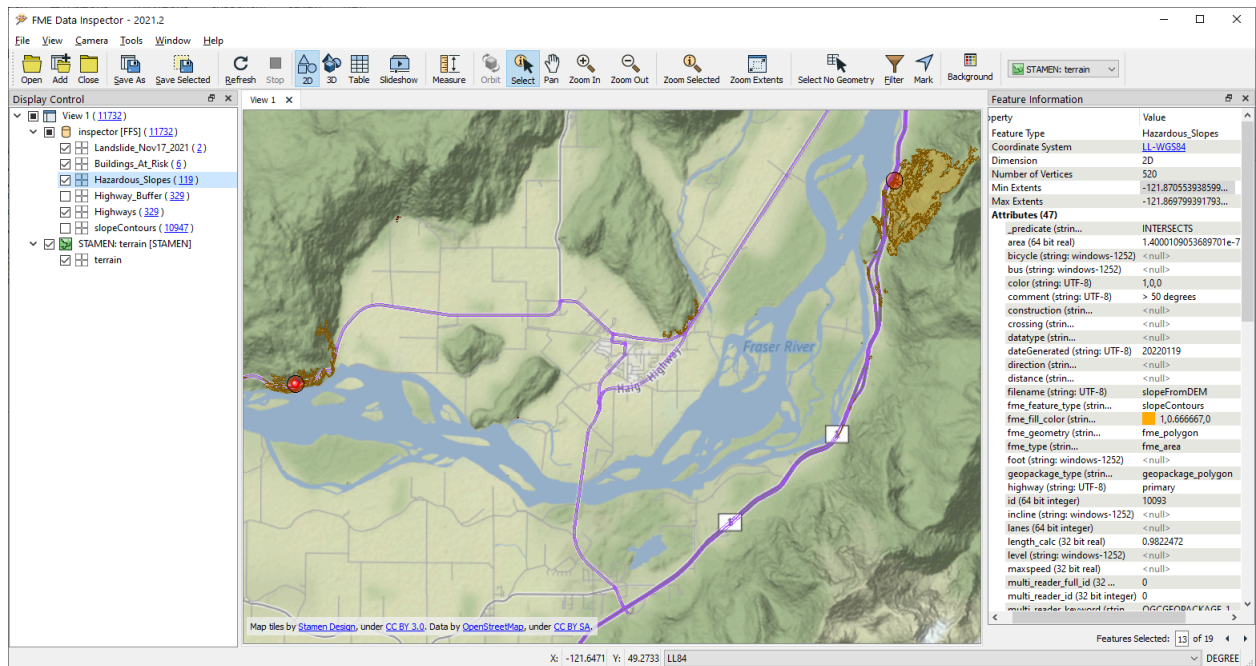
Figure 51

Red circles: Actual landslide locations (approx) from Nov 2021

Orange contours: Steep slopes within 120 m of major highways

In the view above, the areas with very steep slopes are in orange and the approximate actual landslide locations are in red. From this limited example it appears that using the proximity of steep slopes may be a good start in terms of determining potential landslide risks to critical infrastructure.

5.7. Web/Desktop Analysis, Visualization, & Collaboration Component (Safe Software)

5.7.1. Red River Flood Impact Web App

In order to better support end users such as first responders with information essential to their decision making, Safe Software also provided a basic web visualization and reporting application as an example of how DRI results can be distributed to those on the ground during the response effort.

This component takes the ARD and IRD results of the Red River 2011 flood impacts recipe developed by RSS Hydro, DFO, Safe Software, and Skymantics and allows the end user to explore the results by location and infrastructure type.

The source datasets include the Flood Contours Geopackage for 70 timesteps (300K records) and OSM Infrastructure for roads, buildings, and amenities.

An FME workspace was designed to accept these inputs plus a range of user selections and generate a web map of affected infrastructure. User parameters include impact type, flood day, impact severity, minimum flood depth, and buffer amount. The output is an HTML report with Esri Leaflet insert web map. This workflow was published on FMECloud hosted by Amazon AWS, a no-code publication of desktop authored workflow with highly scalable architecture.
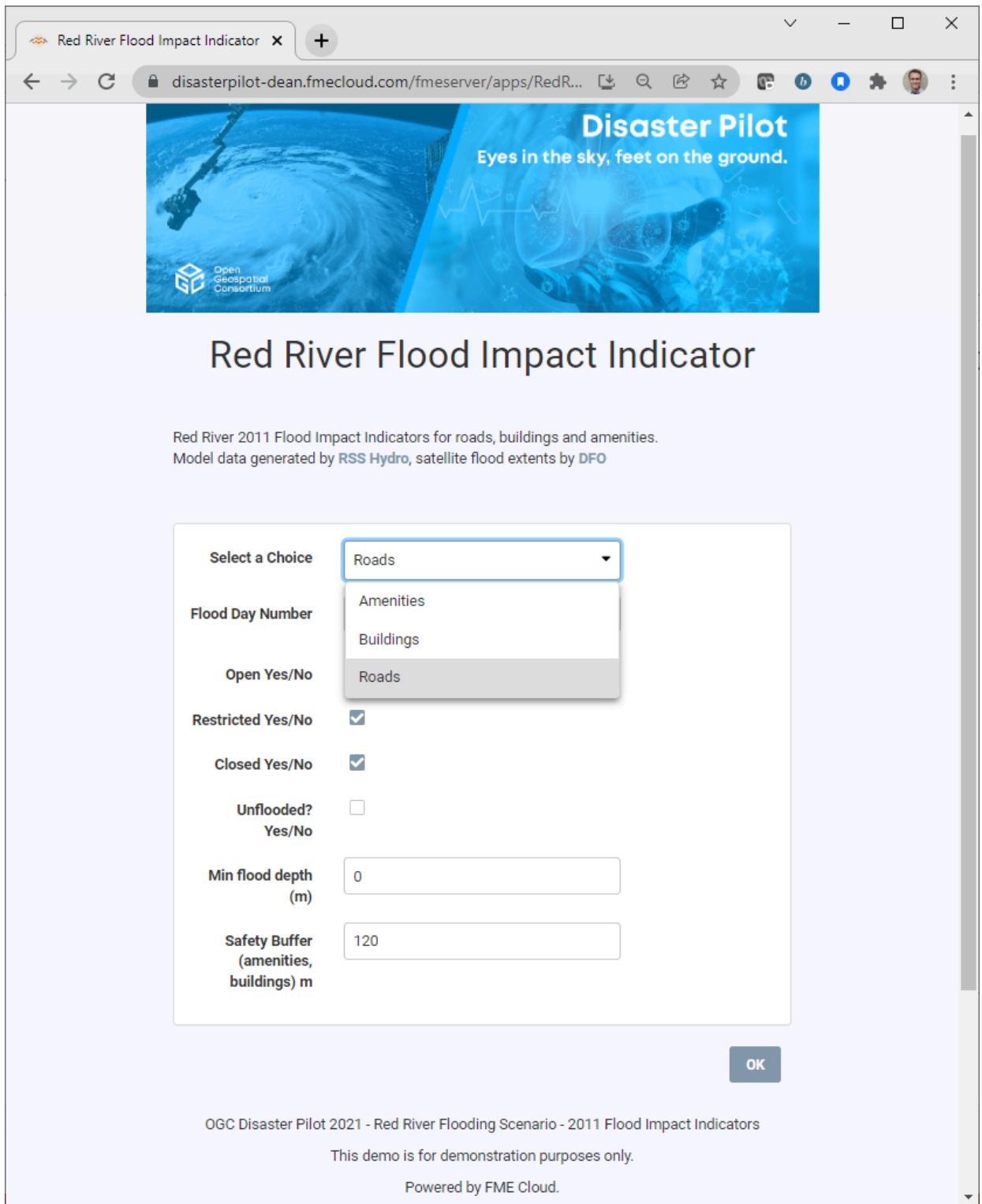
Figure 52

Red River Flood Impact Web App user input form

See: https://disasterpilot-dean.fmecloud.com/fmeserver/apps/RedRiverFloodImpacts
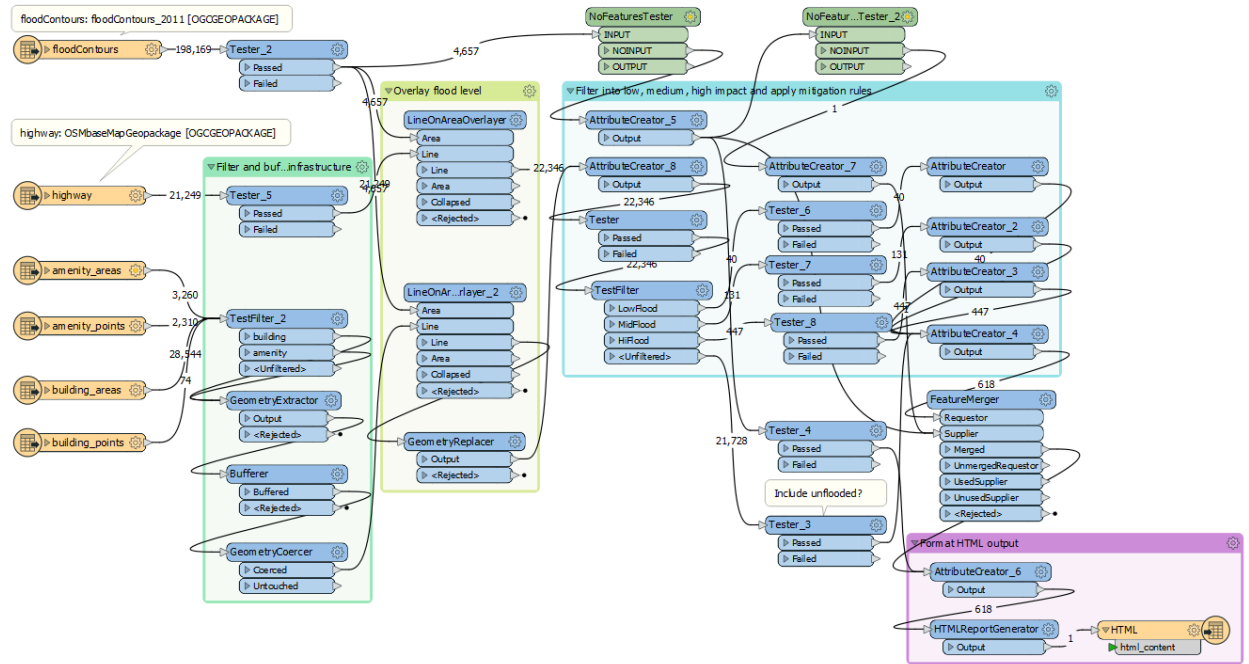
Figure 53

FME workspace — workflow that accepts user inputs, queries the flood contour time series and generates a flood impact reports. This workspace is published to FME cloud as a web service.

Figure 54

Transportation impact report showing flood affected roads including estimated water depth at locations of interest

5.7.2. Technology Interchange Experiments

5.7.2.1. EO Raster — STAC Reader

One of the key data sources for the Disaster Pilot was Earth Observation (EO) data. Due to the large size of the datasets involved and the frequency that they are updated, there has been a lot of effort recently to host EO data in the cloud with cloud native formats and services such as STAC (SpatioTemporal Asset Catalog) and Cloud Optimized Geotiffs (COG). STAC, a relatively new cloud-native format, is a collection of JSON files that create a linked structure for extracting STAC Items. STAC items are typically GeoJSON features that contain links to spatiotemporal data in a cloud environment.

Though STAC is not officially supported by FME, there are built-in transformers that allow FME to interact and access STAC assets. An FME workspace was used to create a custom reader for the format. By executing a series of HTTP GET requests, FME Workbench can filter and import SPOT 4/5 orthoimages of Canada as Cloud Optimized Geotiffs (COG) using the STAC API (v0.8.1). The SPOT 4/5 Collection offers nation-wide EO data from the SPOT Gen4 and Gen5 Satellite's for 2005-2010 and is provided by the Government of Canada.

In a disaster pilot context, FME was able to download and sort each of the 4 raster band files from a SPOT 5 2006 Collection (RGB NIR). The bands are combined using built-in transformers (Sorter/Reproject/RasterBandCombiner) in order create a grayscale raster image for Lake of the Woods, near Winnipeg, Manitoba (North-East of the study area for the Red River Flooding Scenario).

Due to naming convention challenges during the downloading process, there was difficulty producing an RGB image from the STAC resources, so most of the evaluated outputs appeared similar to a false color composite. There were also some limitations to the parameterization of how the STAC reader was called which could be improved on in the future — such as easier to use spatial and temporal queries. Later this process was optimized using FME's S3Connector that allowed for more direct access to the STAC assets. The challenges around collection querying limited the utility of this approach for this phase of the Disaster Pilot. STAC may be revisited in future disaster pilots.

On the data publication side, we ran a number of interoperability tests related to GeoNode.

An upload of a variety of OGC datasets (Geopackage, GeoJSON, and GML) based on OGC standards all failed. The only data format that could be uploaded successfully was Shape and GeoTIFF. A CSW could not be located to automate publication of layer metadata, so this information had to be entered manually each time an upload was processed.

Better success was achieved by accessing services and downloading and reading datasets such as as GML, Geopackage, and GeoJSON from GeoNode using OGC standards for datasets and services. The geometries and coordinate reference systems were downloaded correctly. However, some of the attributions for some of the layers appeared to be missing. On the other

hand, the WFS for the Red River Flood Contours layer seemed to preserve all the feature attribution such as flood depth, time stamp, etc. — see screenshots below.
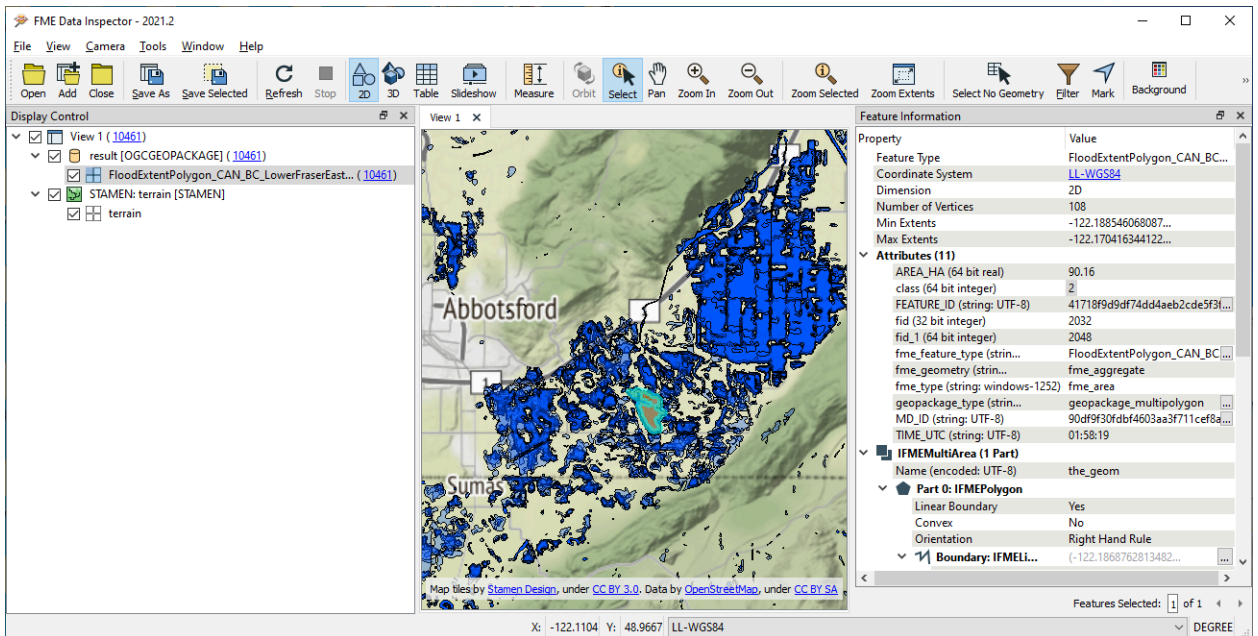
Figure 55

BC flood extent polygons for Fraser Valley East — Geopackage
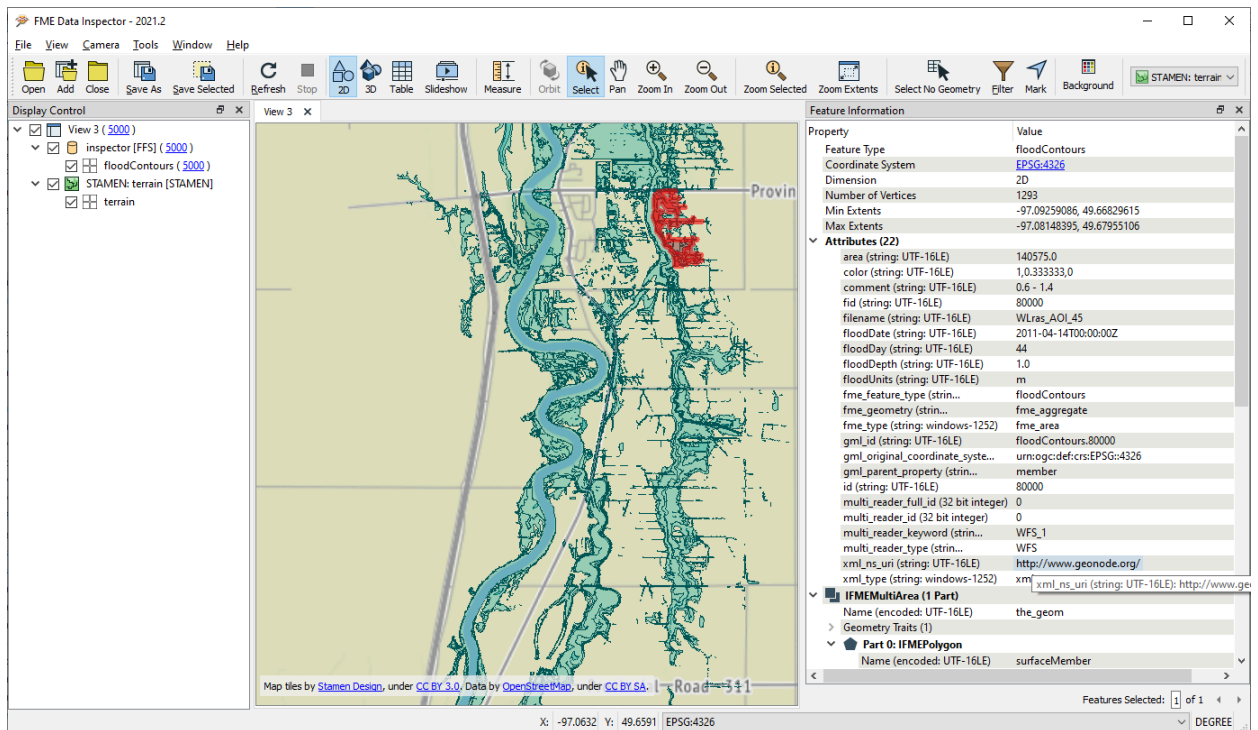
Figure 56

Red River Flood Contours: GeoNode OGC WFS via FME Inspector Client

5.7.3. Technical architecture

Below is an architecture diagram showing where FME components fit into the overall pilot architecture. The FME ARD component consumes EO data and provides ARD datasets to the ARD datastore within GeoNode hosted in the cloud. The FME DRI component takes that ARD and refines it into integration ready datasets (IRD) for use by various analysis and reporting components to support downstream DRI. Finally the FME Web application and reporting component takes a variety of ARD and IRD inputs and DRI results and publishes them in the form of a simple web application that allows end users or disaster responders to interact with the results via a simple web interface.

At each stage the FME components interact with other components via open standards where available. In some cases the source datasets were not in an open standard and had to be converted from proprietary or application specific standards such as OSM or Geodatabase to OGC standards such as Geopackage or other open standards such as GeoJSON.
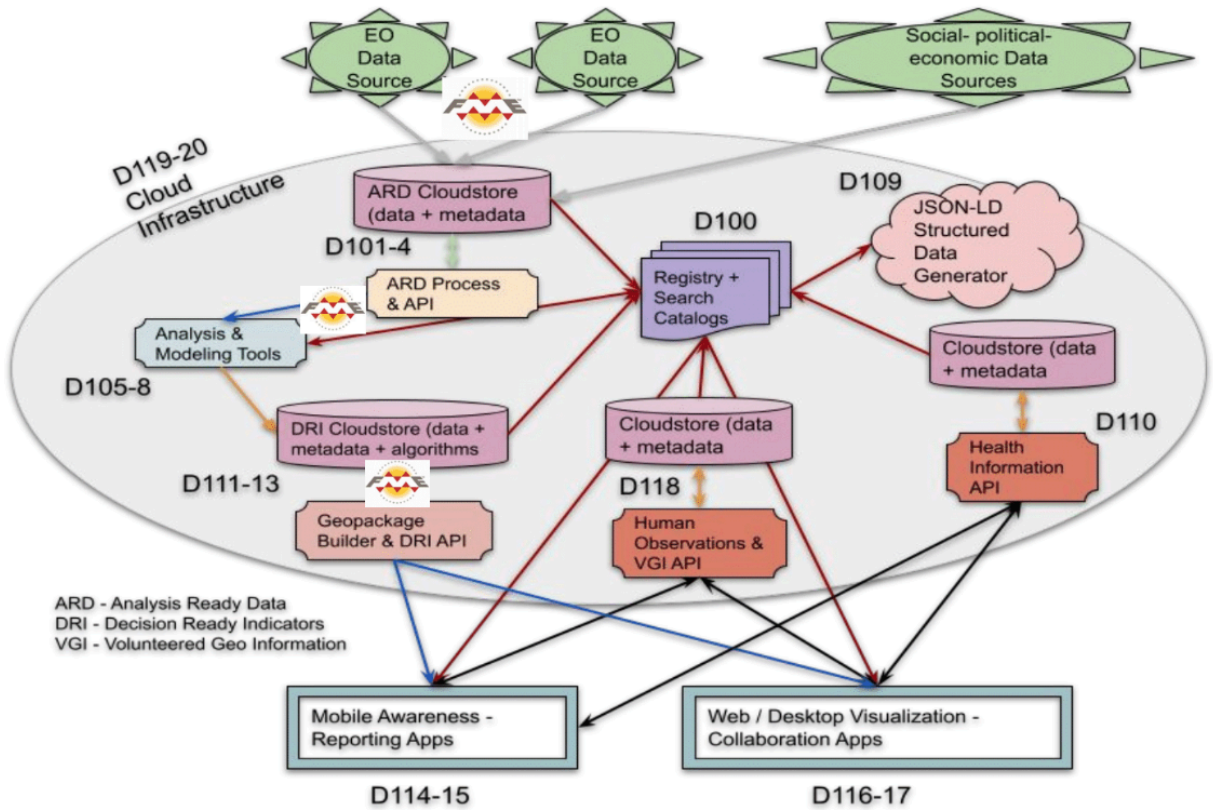
Figure 57

5.7.4. Other Supporting Efforts

Besides contributing the components described above, the Safe Software team also supported the pilot in a number of ways. Safe Software conducted extensive research to explore available

source datasets relevant to the disaster response indicators of interest. This included obtaining base map data from OSM for all the scenarios: Red River Basin, Louisiana, and Peru.

Safe team members also obtained hurricane coastal inundation data from the National Oceanic and Atmospheric Administration (NOAA) for the Gulf Coast, including Louisiana, and did some preliminary coastal inundation modeling and vectorization. However, it was then decided that coastal inundation would not be a part of this phase of the pilot, but may be considered for subsequent phases.

Participants from Safe Software also supported ad hoc data conversion requests for other pilot participants, such as conversion of Geodatabase to Geopackage for those without ArcGIS access, or mosacking and downsampling very large LIDAR based DEMs to smaller, lower resolution DEMs to promote easier access. Safe Software team members also attended many working group meetings in addition to the weekly pilot status calls and made many contributions to in depth pilot design and development discussions.

In addition to working with the 3 pilot scenarios, Safe Software went beyond its commitments to test how adaptable its workflows were by applying them to the BC flooding that happened during the pilot in November, 2021. Recipes for both flood risk and landslides were explored in the BC disaster context. While these were only preliminary efforts, they yielded results that should be useful to inform future test bed activities should they continue to include flood or landslide scenarios.

Finally, Safe Software provided free FME licenses to any participants requesting them for the duration of the pilot. Safe Software also covered the costs of hosting its web application on FME Cloud / Amazon AWS.

5.7.5. Challenges and Lessons Learned

Given the range of scenario types and locations, not to mention the diversity of participants and tools involved, there were clearly significant challenges to implementing a working Disaster Pilot given the compressed timelines involved. On the other hand, these types of demands are not unusual for a typical disaster response effort, and thus even the limitations on success imposed by them can inform some important lessons for future pilots and real world implementations alike.

During the design process it was important to feed back DRI requirements to inform the development of the ARD / IRD component. Initial flood contour output was too coarse to be of use for transportation indicators related to road restrictions and closures. Early output showed water depth in meters, but the indicator required depths in decimeters. For example, the depth of 30cm was determined to be a typical value for closing roads for the public but maintaining access for emergency vehicles. For this reason the ARD component was recalibrated to produce outputs at the decimeter resolution, with contours for 0.1, 0.3, 0.5 meters etc.

The GeoNode hosted by HSR was an invaluable resource and was impressive in its ability to ingest a range of dataset types and dynamically provide OGC services, data downloads, and metadata for them. One problematic limitation was that while there were a number of OGC standards supported for web services and data download, the only uploadable data format was Shape. Attempts to upload Geopackage, GeoJSON, and GML all failed. This should be addressed

in subsequent pilots. In addition, layer metadata had to be entered manually each time an upload was processed because an appropriate CSW could not be located.

A central pilot goal was to provide components that are able to rapidly rerun recipes in near real time. The FME part of the Red River flood impact recipe was shown to be reusable in other contexts: Peru and BC.

On the other hand, in this pilot only a few recipes were developed for impacts ARD and DRI. The main pilot indicators centered around the impact of the natural hazard on transportation. Naturally it would also be important to develop a wider range of indicators related to disaster impacts on populations, built up areas, water quality, food security, and other critical infrastructures such as power and communications. Perhaps subsequent pilots can explore both a wider range of hazard types and a broader diversity of impact types.

The key challenge to running similar workflows for Peru was the limited or low resolution of the accessible DEMs for that region. This did not allow for running of a similar flood model to the Red River scenario. It also meant that meaningful flood depth contours could not to be generated, only flood extents as derived from SAR imagery.

Some hazards are much more difficult to observe / predict at scale, such as landslides. Floods, on the other hand, can be observed and forecast at both large and small scales.

Most scenarios used historical data amd there is the need for more integration with real time flooding services and future forecasts. Methods for local and regional downsampling from global climate models would also benefit modeling of forecasts for future pilots.

Overall, a standards based approach (OGC) was an important part of supporting interoperability and automation. GeoTIFF, GeoJSON, and Geopackage allowed data to be readily exchanged between pilot components. Data uploaded to GeoNode and published to OGC services was easy to explore, download, and access via these services. However, in the context of this phase of the pilot, it was clear that while pilot components were agile and adaptable, most were not fully service connected and required manual data transfers for updates.

Having ready access to both historical and near real time current disaster data allowed users to explore physical and spatial characteristics that may serve as proxies for hazard risk such as height above closest water and steep slopes above near roads. Comparing these characteristics with historical hazard data may reveal patterns that can be used to help estimate future risk and prioritize areas for monitoring.

Even though the scalability and convenience of cloud services were seen as a central pilot goal, most participants experienced difficulties accessing cloud resources for the pilot. Some expected satellite data sources were also not provided in time to be incorporated into the pilot. In the future it would likely be helpful for both cloud infrastructure and data services such as EO to be secured before the pilot is launched.

5.7.6. Next Steps: Updates and Applications

- Build out architecture to use more services / APIs.

- GeoNode should handle open standards inputs.

- Add other rules and analysis: population impacts and evacuation.

- Add other EO & sensor sources for near real-time flooding.

- Integrate with more modeling components to support future projection scenarios (climate change flood model forecasts, down-sampling).

- Review existing data standards and models for disasters (CAP etc), metadata, impacts, and develop new ones as needed.

- Deploy process chain to cloud to ease implementation and improve performance.

- FME enhancements: enhanced support for OGC APIs, both for FME Engine and FME Server / Cloud.

- Future disaster pilots need to take into account more detailed climate model inputs. For looking at climate change effects, the change in past patterns can be examined under climate change by a forecast year given the IPCC climate model projections of factors such as rainfall. This change factor can then be used to proportionally change the flood parameters such as extent and depth.

5.7.7. Summary

The disaster scenario applications discussed above demonstrated that ARD and DRI recipes, indicators, and the associated component inputs / outputs can be used to drive data transform pipeline development. A model based approach proved useful to minimize data wrangling and promote rapid prototyping. Transform models were able to integrate between different datasets and types, automate, and integrate with the cloud. Because available decision rules were constrained by data design and resolution (IRD), feedback from DRI to IRD and ARD was key to refining indicators and increasing their effectiveness. Preprocessing analysis, simplification, and normalization enabled interactive gaming with rules to support multiple indicators.

Finally, utilizing open data standards and common data models was shown to be key to support readiness and collaboration across mixed systems. Ultimately the disaster response scenarios described above illustrate practical examples of how pilot components can be tied together in a data value chain. This value chain showed approaches to extract source data, generate ARD, and refine that as IRD and DRI to support some key disaster response indicators relating to flood and landslides and mitigate impacts on transportation and other infrastructure.

5.8. Analytical Processing Component (StaCen)

5.8.1. Goals of participation

SatCen's participation in the pilot aimed to define and integrate a processing component in a cloud platform near to data so it can be a puzzle piece than help to connect all the players from

data providers to first responders and decision makers. This processing component computes the change detection using Sentinel-1 imagery and is useful for a wide range of hazards like landslides, floods, wildfires, hurricanes, or oil spills. Generic processing was already available, but it has been adapted to the specific scenarios of the pilot.

The SatCen Research, Technology Development and Innovation (RTDI) unit works in the different phases of the data value chain in support of the decision making and actions of the EU external action. SatCen RTDI know-how has demonstrated to be valuable in different phases of the pilot; from providing requirements on Analysis Ready Data (ARD) and Decision Ready Information (DRI) to the internal development of a geospatial platform that provides SatCen with a valuable experience considering Analysis, Visualization, or Cloud Infrastructure components.

Following the successful experience as participants of the OGC EO Application pilot, SatCen participation has allowed to better understand and improve the interdependencies and interfaces between the different components and stakeholders, in order to enhance final users' experience.

RTDI service portfolio includes already several applications of change detection in S1 and S2 imagery, which have been already used in the context of security and disasters (deforestation, illegal mining, wildfires, floods). Such applications could be explored for the scenarios of the Pilot as part of the D105-8 Analytical Processing Components.

5.8.2. Description of the contributed component(s) and role in the Pilot architecture

SatCen has adapted and integrated an application that automatically detects changes using Sentinel-1 (SAR) data, benefiting of the free availability of the Copernicus Programme archive. It computes a change map (that could be used to monitor directly areas affected by floods for example) and other products that can be used later by analysts or other processing components (e.g., RGB composites of amplitude change detection or multitemporal coherence).

Similar applications in the Space & Security domain have already been implemented through R&I Projects by SatCen and exploited for the monitoring of floods (e.g., flooding in Char Piya island in Bangladesh), burned areas (e.g., in H2020 BETTER project) and other man-made and natural disasters (e.g., damage assessment). Those applications are deployed, packed as docker images (with additional metadata describing their data and infrastructure needs) and registered in the internal SatCen platform named GEO-DAMP.

GEO-DAMP (Geospatial Data Management Platform) is a Web-based Platform for the management of geospatial products and services, designed following a modular microservices architecture. The Platform conveys a collaborative environment where data can be discovered (through a CS-W and OpenSearch interface), accessed for visualization and download (through a WCS interface), shared and processed. The Platform provides a common environment within SatCen for the integration and validation of new applications facilitating their transition into operational scenarios.

Similar applications were already integrated in different platform providers during the OGC EO Applications pilot, where relevant conclusions and recommendations were produced. Based on

this experience, SatCen will help to assess the validity of these recommendations in the disaster pilot.

5.8.3. Technical design and implementation

5.8.3.1. GEO-DAMP platform

The adaptation of existing pipelines has been carried out in the geospatial platform GEO-DAMP. This platform was put in place by SatCen RTDI unit in order to demonstrate the potential of working in a platform-based approach.

GEO-DAMP makes use of innovative conceptual (e.g., continuous improvement and value delivery), architectural (e.g., cloud native micro-services pattern), and technical (e.g., EO data cubes, distributed processing) solutions. It constitutes an advanced environment where EO users and EO application developers can interact for continuously enhancing capabilities in the Space and Security domain, enabling to cope with the growing demands throughout the different phases of the data value chain (e.g. discovery, access, processing and exploitation of EO and collateral data). In addition, GEO-DAMP allows to share results and knowledge, opening the door for collaboration between relevant stakeholders.

GEO-DAMP, based on common standards (e.g., OpenSearch, WMS, WFS, WCS), gives to its users access to different datasets (EO and not EO) in a clear way from a single platform. At present, users can access data from Sentinels (e.g., Sentinel-1, Sentinel-2, Sentinel-5P), Landsat 8 and other data sources (e.g., Automatic Identification System). SatCen is continuously working towards the integration of new relevant EO (e.g., Very High Resolution Data) and non-EO (e.g., Open Source, Citizen Science) data sources.

With regard to data processing, the integration of new pipelines is standardized based on containerization technologies. EO application developers can provide arbitrary complex cross-platform applications simply by providing a docker container and supportive metadata about the process, and GEO-DAMP will orchestrate them accordingly to the underlying hardware infrastructure available and processing needs. GEO-DAMP makes use of efficient cloud-native technologies, logically built as a set of decoupled services packaged in containers and supported by modern orchestration technologies (k8s) to ensure its scalability, flexibility and cloud-agnostic capabilities.

Moreover, GEO-DAMP provides SatCen with the possibility to exploit new services relevant for Security based on the available datasets: Amplitude and Coherence Change Detection using Sentinel-1 data as well as Change Detection, Characterisation, Thematic Indexes, and Time Series using Sentinel-2 data are currently accessible through GEO-DAMP and being used to monitor several areas of interest.

GEO-DAMP strengthens SatCen capabilities to discover, access, process, and exploit different datasets, as well as unlocks the potentiality for collaboration in the Space and Security domain. Furthermore, GEO-DAMP is guaranteeing the sustainability of the results obtained from the different R&I projects implemented by SatCen.

GEO-DAMP is being continuously improved by seeking, implementing and validating the use of innovative solutions along the complete data value chain in order to maintain SatCen capabilities

most up to date and to be able to efficiently cope with the growing security challenges of our times.
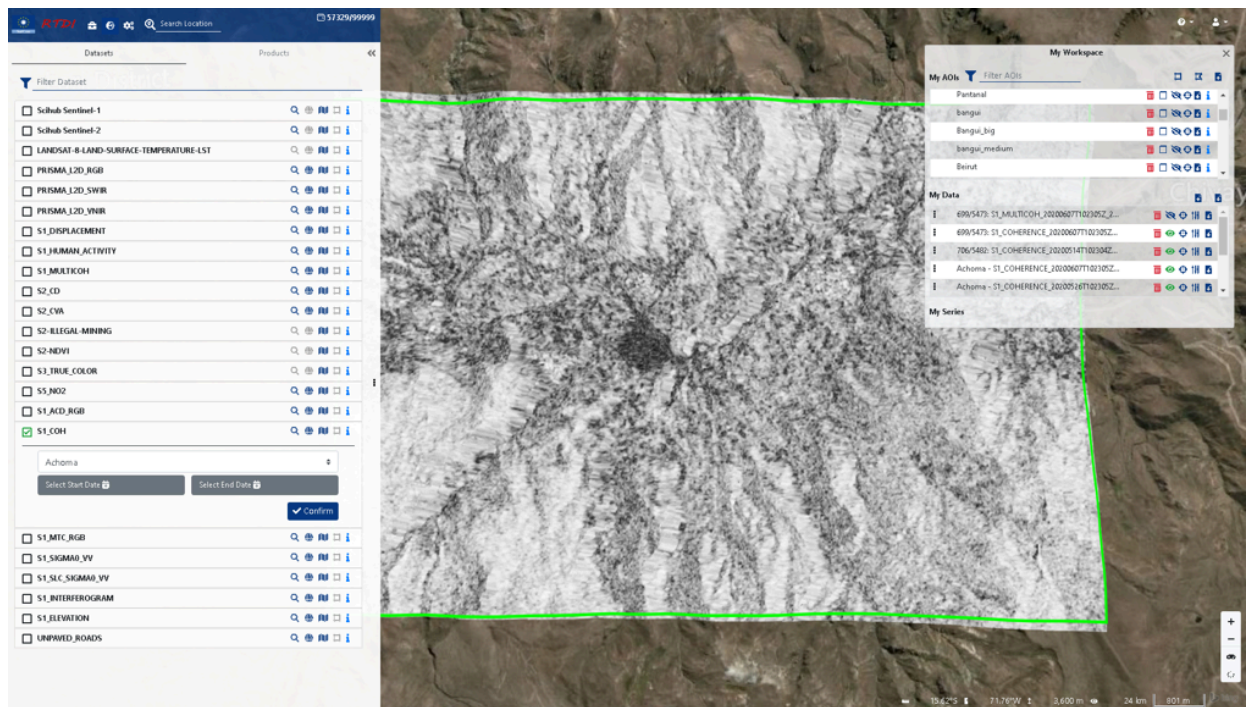
Figure 58 — GEO-DAMP Platform

5.8.3.2. Processing pipelines

The processing pipelines developed and adapted for the pilot take advantage of Copernicus data. Copernicus is the EU Earth Observation programme and offers services that draw from EO and in-situ data. In particular, the imagery from the Sentinel missions that are part of the Copernicus space component has been exploited.

The focus has been put in the missions Sentinel-1 and Sentinel-2.

The Sentinel-1 mission includes a Synthetic Aperture Radar (SAR) sensor that can work in different imaging modes with resolutions from 5m and coverage up to 400 km. It provides dual polarization capability and short revisit time (6 days considering satellites Sentinel-1A and -1B). This active sensor allows the collection of useful data even when clouds are present, since it is able to penetrate clouds.

The Sentinel-2 mission includes a multispectral sensor with 13 spectral bands, swath of 290 km and resolutions up to 10 m. The revisit time is 5 days when considering both Sentinel-2 satellites. The main disadvantage of this sensor is that is affected by clouds, which reduces the usability in cloudy areas or during some events characterized by the presence of clouds (e.g., flooding caused by rainfall).

5.8.3.3. Amplitude Change Detection (ACD)

ACD relies upon the combination of the "Amplitude Information" from two SAR images into a single image. ACD allows the comparison of the backscatter of two images. ACD should be computed using as input images acquired in the same conditions (acquisition mode, relative orbit) and corrected both geometrically and radiometrically.

This product is very sensitive to changes that affect the backscatter like:

- New buildings: before the new construction there is low backscatter and after, high backscatter;

- Flooding: there is a decrease of the backscatter because of very low backscatter of water bodies; and

- Burned areas: depending on the vegetation characteristics, there could be an increase or a decrease of the backscatter.
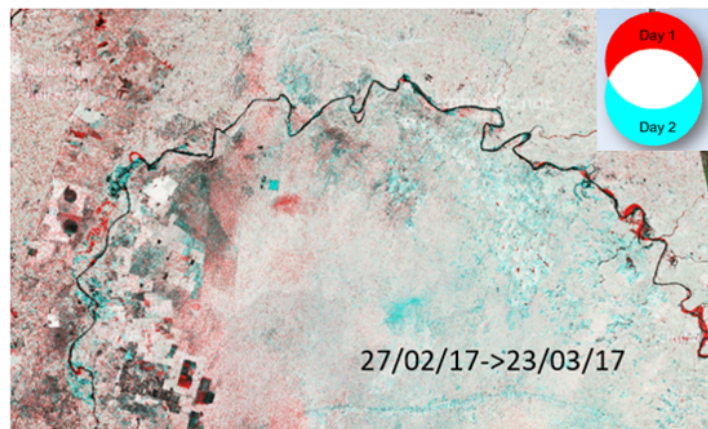
Figure 59 — Amplitude Change Detection example

5.8.3.4. Multi-Temporal Coherence (MTC)

SAR images consist of amplitude and phase components: the amplitude reflects the intensity of the backscatter while phase is related to the distance of the target and its characteristics.

It is possible to have a measure of the phase stability between two acquisitions by computing the coherence.

While changes that causes differences in the amplitude can be distinguished in the ACD, changes with a low change in the backscatter, can be detected with the MTC. For example, typical applications of MTC are landslides (that usually produce changes in the phase but not in the backscatter amplitude) or changes in artificial structures.
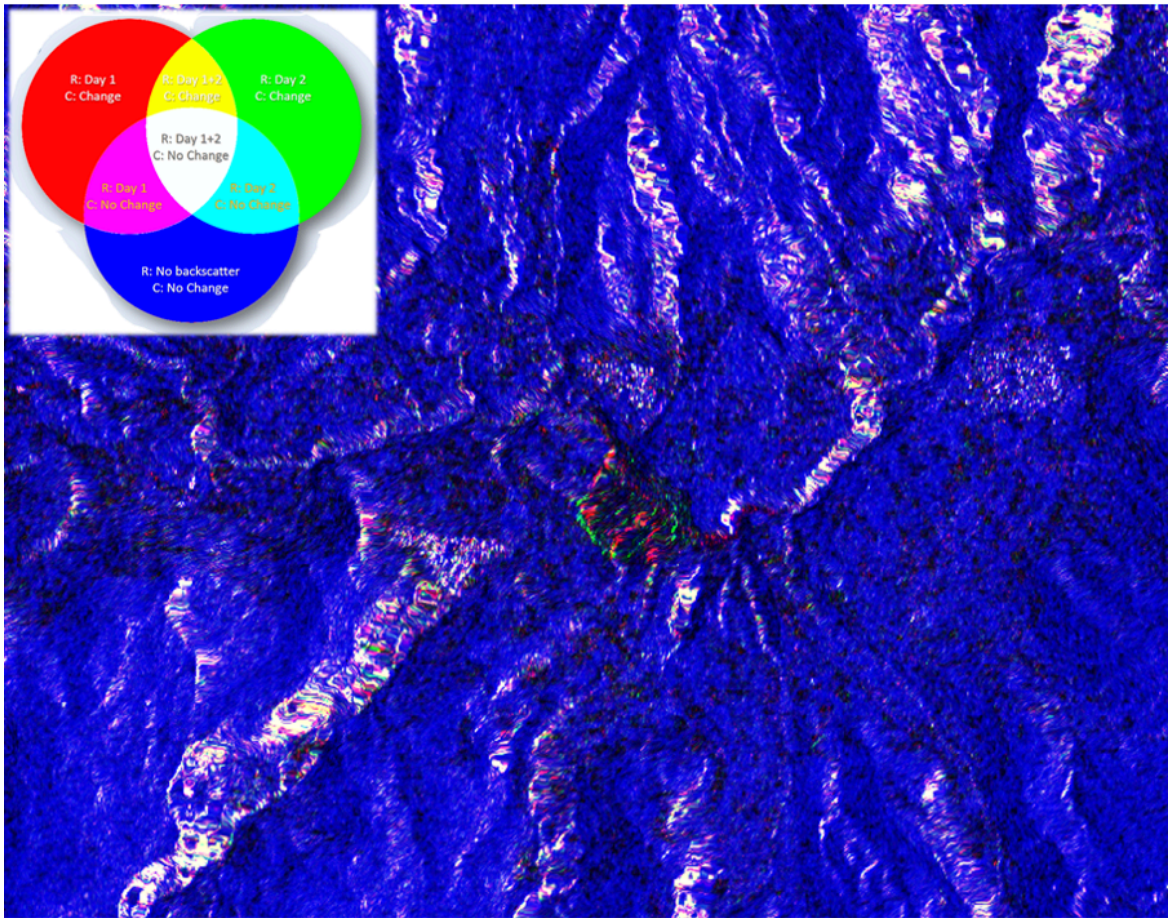
Figure 60 — Multi-Temporal Coherence example

5.8.4. Scenario workflows and recipes

The technical solutions described previously have been exploited for two different disaster scenarios: flooding and landslides in Peru.

In these scenarios, remote sensing data from space is very useful for the characterization and monitoring of large-scale phenomena such as floods, as it allows users to obtain data over large areas at a scale difficult to reach using field-based instruments and methods.

5.8.4.1. Flooding

For flooding, two main sensors have been considered in the analysis.

- Sentinel-1 — SAR: SAR is very useful for mapping flood extent since it can acquire images in all weather conditions. However, its adequacy depends on the characteristics of the

area under analysis. In this sense, different strategies have to be applied depending on the characteristics of the terrain.

- In **open areas**, water surfaces are smooth and the specular reflection produce low backscatter (black pixels in the image);

- In **forested areas**, if the SAR penetrates the canopy, the backscatter is higher than the reference image in flooded areas due to double bouncing; and

- In **urban areas**, due to the strong scatter, it is difficult to detect flooding with SAR.

- Sentinel-2 — Optical: Optical sensors can also be used for mapping flood extent. The changes are easily detected visually and algorithms like Change Vector Analysis can be applied to automate the task. The main disadvantage of optical data sources is their dependency on weather conditions since if there are clouds, no information is available.

Based on the above-mentioned sensors and solutions, it has been possible to extract the flood mask through image segmentation techniques such as simple thresholding. The flood mask can be consequently used to monitor the extension of the areas affected, as well as to overlay it with reference maps (e.g., as obtainable from Open Street Map) to identify possible affected critical infrastructures.

5.8.4.2. Sentinel-1 Change map recipe

The inputs for this recipe are a couple of Sentinel-1 GRD products. The process consists on the application of ACD algorithms to generate as output a COG (Cloud Optimized GeoTIFF) that can be used easily by final users to visually identify changes. The changes are highlighted in red and cyan. See example in Figure 2.

The outputs can be converted to byte in order to facilitate the visualization or can keep the calibrated backscatter as float32 to facilitate their use as ARD (Analysis Ready Data) by other processing chains.

5.8.4.3. Flood mask recipe

The input for this recipe can be a Sentinel-1 GRD product or an ACD. The process consists of applying a threshold to the backscatter (or to an index computed with the backscatter at two different dates) to obtain a binary classification of the image: flooded/not flooded pixels. The output is generated in both raster (COG) and vector (GeoJSON) format.

The output is the flood mask, that can be considered as DRI but also as input for other recipes (e.g., 'intersect' the flood mask with roads to obtain affected infrastructures or to compute alternative routes).
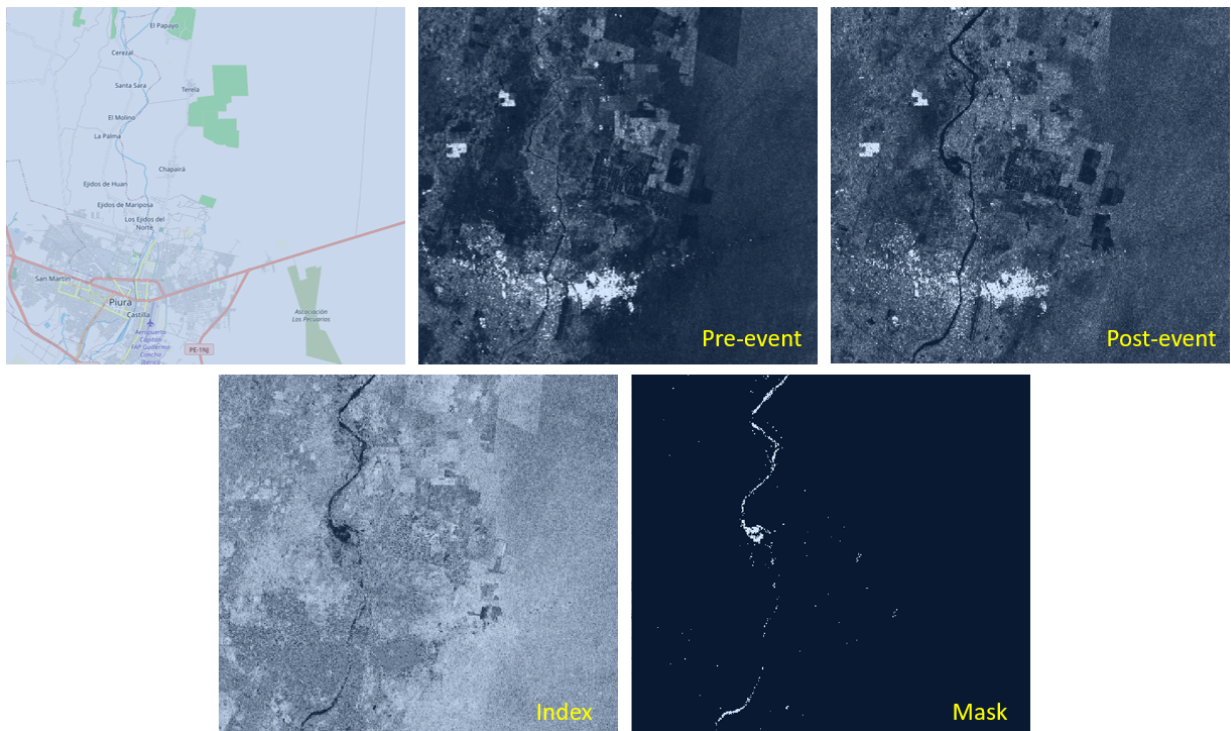
Figure 61 — Example of flood mask computed from Sentinel-1 data; generated by SatCen.

5.8.4.4. Landslides

The main event studied during the pilot is the Achoma landslide that occurred on 18th June 2020. In this event, soil and rock on a hillside slipped loose and created a landslide affecting more than 40 hectares. The landslide generated a dam in the Colca River, which caused flooding. The event is clearly visible using Sentinel-2 data, as shown in the figure below.
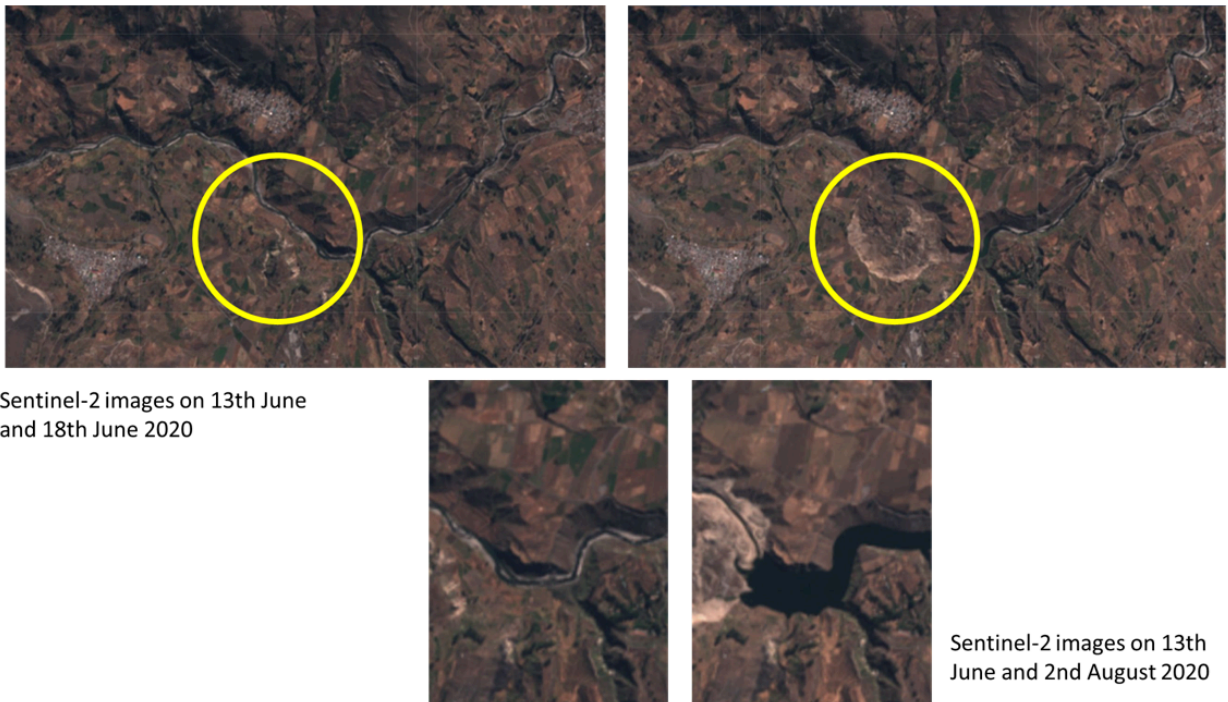
Sentinel-2 images on 13th June
and 18th June 2020

Sentinel-2 images on 13th
June and 2nd August 2020

Figure 62 — Sentinel-2 images over Achoma landslide; area of the
landslide is highlighed with the yellow circles; generated by SatCen.

5.8.4.5. Sentinel-1 Change map recipe

The inputs for this recipe are Sentinel-1 SLC products. The process consists on the application of Multi-Temporal Coherence (MTC) algorithms to generate as output a COG that can be used easily by final users to visually identify changes. The changes are highlighted in red, green, and yellow.

In this case, the MTC was selected instead of the ACD since although landslides affects to SAR backscatter, it is the coherence (the blue component in MTC composites) where changes are better distinguished.

The RGB output can be converted to byte in order to facilitate the visualization or can keep the calibrated backscatters and coherence as float32 to facilitate its use as ARD (Analysis Ready Data) by other processes.

Another useful output is the coherence (as a separated output, since it is one of the channels of the composite), which is a measure of similarity of radar backscatter, taking into account not only the amplitude but also the phase.

The use of time series of coherence seems to be useful for the identification/monitoring of affected areas but also for the prediction in some cases. In the study carried out, the area affected by Achoma landslide presents an anomaly in the coherence some days before the landslide happened, as visible in the figure below.
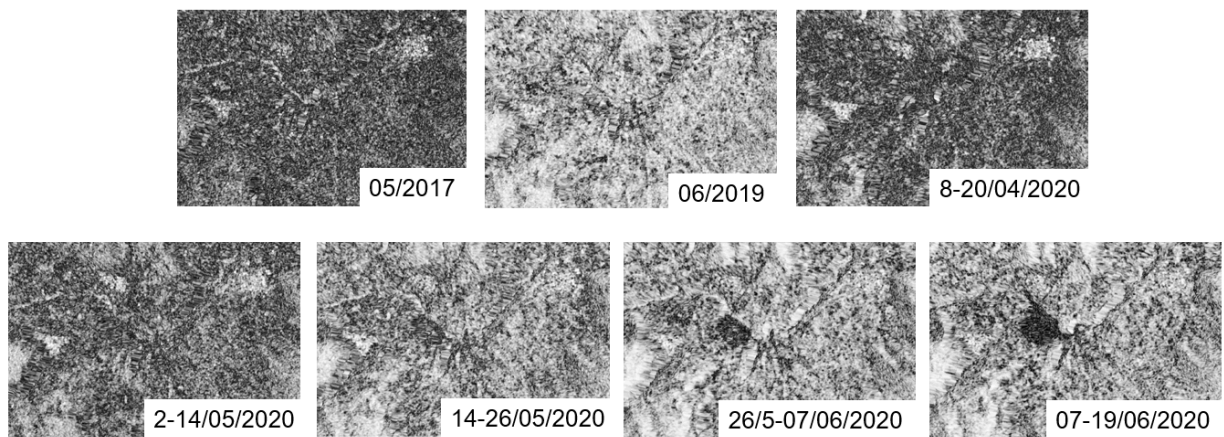
Figure 63 — Time series of coherences generated with Sentinel-1 data over Achoma; generated by SatCen.

In the time series, the coherence in all the area seems uniform. It is higher or lower depending on the specific pair of images (probably to soil moisture), but always homogeneous.

After the pair from 14th and 26th of May 2020, the area where the landslide will happen can be distinguished because of the loss of coherence. The loss of coherence is maximum in the last image (computed with a pair of images just before and after the landslide), and the contour of the affected terrain can be delineated easily (e.g., using the recipe described below).

5.8.4.6. Landslide mask recipe

The input for this recipe is the output of Sentinel-1 Change map (including the coherence). The process consists of applying a threshold to the coherence (using time series information to detect anomalies) to obtain a binary classification of the image: affected by landslide/not affected by landslide. The output is generated in both raster (COG) and vector (GeoJSON) format.

The output is the landslide mask, that can be considered as DRI but also as input for other recipes (e.g., 'intersect' the landslide mask with roads to obtain affected infrastructures).

5.8.5. Interoperability and Technology interchange experiments

The processes have been run in GEO-DAMP platform, which makes use of several OGC standards along the whole data lifecycle.

- Opensearch is used for discovering products to be used as inputs to the processing. Furthermore, it is used to feed the internal catalog hosting the processed outputs.

- Access to the data for processing is centralized in an EO Data Access Gateway component, that works with common authentication schemes (e.g., basic http authorization, based on cookies, tokens, etc.) and standards (OData, S3, etc.)

- The processing itself is implemented in line with the EO Apps to the Data Guidelines, as SatCen has actively participated to the EO Apps Pilot and is following closely the development of an EO Apps initiative.

- Output data resulting from the processing chains is typically published as COG and made available also through a WCS and WMS endpoint.

Different limitations have made it complex to further pursue interoperability experiments. In particular, due to time constraints and the late availability during the pilot of data and credits, SatCen been mostly making use of our internal infrastructure. It is also important to mention that the bespoke algorithm makes use of S1 SLC data, which availability is more limited in partners/sponsors infrastructure as opposed to S1 GRD data.

5.8.6. Challenges and lessons learned

- ARD and Decision Ready Information/Indicators (DRI) are usually difficult to classify. Some ARD could be used directly by decision makers and also products what we are considering DRI, are used as inputs for other applications, so from their point of view they are ARD. Probably any recommendation about standards, formats, etc. for either should be applied for both.

- Sentinel data has demonstrated to be very useful in flooding and landslides scenarios, but its resolution does not allow to cover all cases an affected areas. For example, the usability of Sentinel-1 data to monitor flooded urban areas seems limited.

- Sentinel-2 data can complement very well Sentinel-1 in non-cloudy areas.

- Flooding and landslide disasters can occur by many different reasons. For that, the indicators to monitor the risk and try to predict them have to be adapted for the different specific types of events. The indicators and recipes for the monitoring once the disaster has happened are more generic and valid for most of the cases.

- The analysis carried out in Achoma landslide has shown that the use of time series of coherence could allow the prediction of this type of event at least in some specific conditions, but a deeper study is needed to be able to understand in which scenarios this is possible and with which limitations.

5.8.7. Next steps: updates and applications

SatCen believes there is potential to carry on further interoperability experiments in the future. In particular, investigation has occurred on getting the input data from ARD providers and advertising SatCen desired outputs within the DRI WG for some partners willing to apply post-processing. Also, SatCen's pipelines follow the EO Apps to the data specification, which make them suitable to run across multiple cloud infrastructures (provided the relevant ADES/EMS components are available), as demonstrated through SatCen participation in the pilot last year. Those are definitely topics that should be further explored in future initiatives.

Regarding the applications, the landslide scenario would require more tests in other areas and with other techniques (e.g., interferometry) to try to understand in which cases and what types of landslides could be predicted. Also, the identification of risky areas could be the starting point for a systematic monitoring.

5.9. EO Data Component (SATELLOGIC)

5.9.1. Requirement in satellite data

The EO data used in the Pilot is provided from satellites orbiting the planet. There are different types of satellite data, which are listed below, but two important aspects of all satellite data are their spatial and temporal resolution.

Spatial Resolution can be used to both describe the size of the smallest object that can be seen in an image and the distance on the ground each pixel on an image represents — a 15-meter spatial resolution means that each pixel represents 15-meters on the ground and, in general, objects smaller than 15-meters cannot be seen. For a disaster scenario, a 15-meter resolution might be acceptable to investigate how far flooding has spread across a broad region, but not for determining if a specific road has been flooded.

Temporal resolution is the frequency of the data collection over a specific point on the Earth. Most satellites orbit the Earth and so can only see a part of the Earth at any one time, and can take hours or days to come back to the same point on the planet. This is a challenge with a fast-changing disaster situation. To resolve this issue, some geostationary satellites stay over the same point on the Earth at all times – although data from this type of satellite were not used within the Pilot. Alternatively, some providers use multiple satellites operating as a constellation which mean different satellites image the same area more frequently. Finally, it is possible to use datasets from different satellites to increase the frequency of monitoring; using this approach reinforces the need to have implemented data standards to ensure that the use, integration and comparison of different datasets are simple.

There are a number of different types of data provided by satellites and the most common are the following.

- Multispectral Optical Data is an image of the Earth taken by a sensor onboard a satellite, and the imagery is similar to how the human eye sees the world. The biggest challenge with optical sensors is that they cannot see through the clouds. Example satellites that offer optical imagery include NASA's Landsat missions, European Space Agency's (ESA) Copernicus Sentinel-2 satellites, Japan Aerospace Exploration Agency's (JAXA) ALOS-3, PeruSAT, Planet's constellations & Satellogic's Newsat constellation.

- Hyperspectral Optical Data is collected across a wider part of the electromagnetic spectrum from the visible to shortwave infrared, and these sensors collect lots of individual measurements each of which is a potential dataset. This allows this data to identify and specify features in the land and the atmosphere. Hyperspectral data could be used to identify potential pollutants in the air for disasters, although it has not been

used specifically within this Pilot. Examples offering this type of data include ESA's CHRIS-PROBA for the land and the TROPMI instrument on Sentinel-5P for the atmosphere.

- Microwave Data is the companion to optical data and is captured from the microwave part of the electromagnetic spectrum. The most common type of microwave data is Synthetic Aperture Radar (SAR) data, which has the advantage over optical data in that it can see through clouds and acquire data at night. Examples offering SAR imagery include Canada's RADARSAT, ESA's Sentinel-1, JAXA's ALOS PALSAR and commercial missions such as the ICEYE constellation.

It is acknowledged that EO data is not only available from satellites as it can also be supplied from both piloted aircraft and drones; however, the Pilot only used satellite EO. Aircraft are commonly flown in disaster scenarios, although poor weather or no-fly zones can restrict their use. Remotely piloted drones are potentially a very useful development for data collection, however, currently, they are still relatively new technology. While some government agencies have drones, there are also a lot of volunteer/amateur drone pilots available. These are not currently fully utilized, and solutions need to be sought to marshal such resources such that they complement and do not hinder official drones and agreeing on processes for making the data available.

In addition, EO satellite technology is constantly developing with more satellites being launched offering more data more frequently, together with new technology such as video-based EO. As such, the capabilities of what EO can offer will continue to develop in the coming years.

The sources of data were both free-to-access and paid-for and the availability could be restricted depending on the source and sensitivity of the data. Sources considered or used within the Pilot included the following.

- Satellite Earth Observation:

  - Copernicus Sentinel Missions: Operated by the European Union, with data acquired by constellations of global missions focused on specific technologies. For example, Copernicus Sentinel-1 are two Radar missions (A&B) that operate at the C-band frequency.

  - Landsat-8: A high resolution (30 m spatial resolution) mission that carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments. Landsat 9 is the most recently launched Landsat satellite, but was not launched in time for the Pilot.

  - NewSat: Satellogic constellation currently consists of 17 commercial NewSat satellites in sun-synchronous Low Earth Orbit (LEO).

  - PeruSat-1: A small (~ 430 kg) satellite launched in September 2016, which carries NAOMI (New AstroSat Optical Modular Instrument) that has a panchromatic and four multispectral wavebands: Blue, green red, and Near-InfraRed (NIR). The panchromatic band has a spatial resolution of 0.7 at nadir, while the multispectral bands have a spatial resolution of 2.5 m at nadir.

- RADARSAT: The three RCM (RADARSAT Constellation Mission) satellites operate at the C-band frequency, with a spatial resolution from 1 to 100 m depending on the acquisition mode.

5.9.2. Indicators

Although recipes and indicators vary from disaster to disaster, they can be set up within a structure that includes their applicability timescale (short-term predictions and impacts to medium and long-term predictions) and type/geospatial extent of the disaster. Example recipes include those related to the following

- Predicting future flooding: Using Earth Observation (EO) and modeled data to provide the current ocean state and weather conditions in support of predicting the onset of flooding due to heavy rainfall.

- The blockage of roads by floodwater: EO and modeling data is used to extract/predict the floodwater extent, and then used to determine which and to what depth roads are affected, which is then used to influence the routing of traffic.

- Availability of health supplies: Use of geospatial health data to predict a Pandemic Mortality Risk Index and Medical Supply Needs Index.

5.9.3. Collections

- EO data: Sentinel-1, Landsat-8, NewSat Level-1 from Satellogic, and RADARSAT from NRCan

- Flood inundation area extraction based on EO images

- Population Density data

- Land Cover (10m and 30m)

- 30m SRTM DEM

5.9.4. How to get data

```
Datasets and Data Sources: * EO data
```

Figure 64

Table 2 — EO data

DATA	PROVIDER	SCENES IN RED RIVER	SCENES IN PERU	RESOLUTION
RadarSat	NRCan	10	0	2m
PeruSat	CONIDA	Unknown	Unknown	Unknown
NewSat	Satellogic	9	12	1m
Sentinel-1	ESA	72	10	10m
Landsat-8	USGS	79	28	30m

5.9.5. How to be improve the access

Further work needs to be taken to improve the technical setup for access, e.g., APIs or other computer-to-computer interfaces need to be set up rather than manually transferring data from one provider to another. This approach was being accomplished towards the end of the Pilot and further work is needed for the ecosystem of providers to function smoothly and handle near-real-time data. Then, large scale disaster response test events need to be conducted to stress-test the setup and see where the points of weakness are.

5.9.6. Needs for cover area and time

- Satellites and Sensors

- Satellites Characteristics: Orbits and Swaths

- Spatial Resolution, Pixel Size, and Scale

- Spectral Resolution

- Radiometric Resolution

- Temporal Resolution

5.9.7. Satellite base data to help

Table 3 — Precipitation

DATA	PROVIDER	TEMPORAL RESOLUTION	SPATIAL RESOLUTION
GPM	NASA GES DISC	Daily/Hourly	0.1 degree

Table 4 — Surface Water

DATA	SOURCE	RESOLUTION
Ocurrence	EU, JRC	30m
Ocurrence change intensity	EU, JRC	30m
Seasonality	EU, JRC	30m
Recurrence	EU, JRC	30m
Transitions	EU, JRC	30m
Maximum Water Extent	EU, JRC	30m

The European Commission's Joint Research Centre developed the Global Surface Water dataset (see https://global-surface-water.appspot.com/#features++) in the framework of the Copernicus Programme.

Table 5 — Other Background Data

DATA	SOURCE	RESOLUTION
DEM	SRTM	30m
Population	WorldPop	90m
LandCover	GLC	10/30m
Road Network	OSM	N/A

Table 6 — Derived Ingredient

DATA	SOURCE
NDWI	Landsat-8
NDvi	NewSat
Event Water Areas	EO Data and Terrain Index
Flood Areas	EO Data and Terrain Index
Flood Duration	Flood Areas
Height Above Nearest Drainage (HAND)	DEM
Slope	DEM

5.10. Secure GeoPackage Service Component (Secure Dimensions)

NOTE This section describes the Secure Dimensions contribution Secure GeoPackage Service.

Data Centric Security is an integral part in the Disaster Pilot 2021. This section provides an overview to all the participating components as well as their interoperability necessary to create and use Secure GeoPackage(s).

5.10.1. Introduction

Data Centric Security (DCS) is a design, implementation, and deployment pattern that introduces the concept of data + security to be an integral, atomic compound. In a nutshell, DCS does not rely on any communication security mechanism, such as HTTP over TLS (HTTPS) or VPN, etc. Instead, the concept introduces encryption of the data and the associated metadata to enable decryption by legitimate users and trusted applications. Applying DCS ensures confidentiality at all times: while the data is at rest (i.e., stored on a cloud drive, on a desktop, or mobile device) and in transit (sent from one machine or device to another regardless of communication security). This concept is extremely important and beneficiary when protecting sensitive information in situations where storage security capabilities are unclear. One example is cloud storage where providers may decide (upon usage matrices) if a storage container is moved from a high-speed segment to another, slower and perhaps less secure environment. In some cases this change of storage might include change of physical locations, even storage facilities across country borders.

The use of encryption to protect sensitive data towards confidentiality does imply that the management of the (decryption) keys is thought out properly: If access to the keys is possible for unauthorized users or malicious applications, the confidentiality is at risk. Therefore, the operation of a "fit-for-purpose" Key Management System (KMS) is a paramount requirement. To prevent shortcomings and oversights when designing, implementing and deploying a Key Management Server, the NIST 800-57 Recommendation for Key Management: Part 1 – General is an extremely important publication.

The NIST 800-57 publication introduces the life cycle of a key from its creation until the end of life (that actually is never reached if data was encrypted with the key) and the aspects of key protection; most important access control. Therefore, an integral part of a KMS must be the ability to allow admins and key owners to manage access conditions to the key(s) that respect the life cycle phase of the key but also other aspects like sharing keys with other users and applications. This aspect is very important for Disaster Management, as encrypted data might be shared with first responders and other involved parties. Typically, the access conditions are based on:

- the identity of the consuming user (either based on identity UUID or email address);

- the identity of applications (to ensure they are trusted);

- time (what is the start and end times that a key can be obtained); and

- location (where must the user's device be located to enable access to a key).

For this OGC Pilot, a system was put in place that allows to create GeoPackages with encrypted data. An authorized user can request an encrypted GeoPackage via the Decision Ready Information Component (D111-3) — also known as the GeoPackage Encryption Service (GES) — that implements DCS capabilities on top of the OGC Web Map and Web Feature Service protocol. For each request, a unique encryption key is used that was either generated before the request or generated when processing the request. Either way, the key must have been registered with the KMS before making a request to the GES. With the response, the GeoPackage with encrypted content, a reference to the key is added to the GeoPackage. Assuming the application (and the acting user) have the appropriate access rights on the key, the KMS releases the key and the application can decrypt the content.

The following section introduces an infrastructure and their components that support the creation, sharing, and use of GeoPackages with encrypted content in this Disaster Pilot. The introduced architecture (see Figure 65) has mainly being developed throughout OGC Testbeds 15, 16, and 17 within the Data Centric Security thread. The component Authorization Server was developed, implemented, and is operated by Secure Dimensions outside of OGC initiatives. The DRI Component D111-3 and the structure of the GeoPackage with Encrypted content (see annex Annex A) has been developed, implemented and deployed within this Pilot.
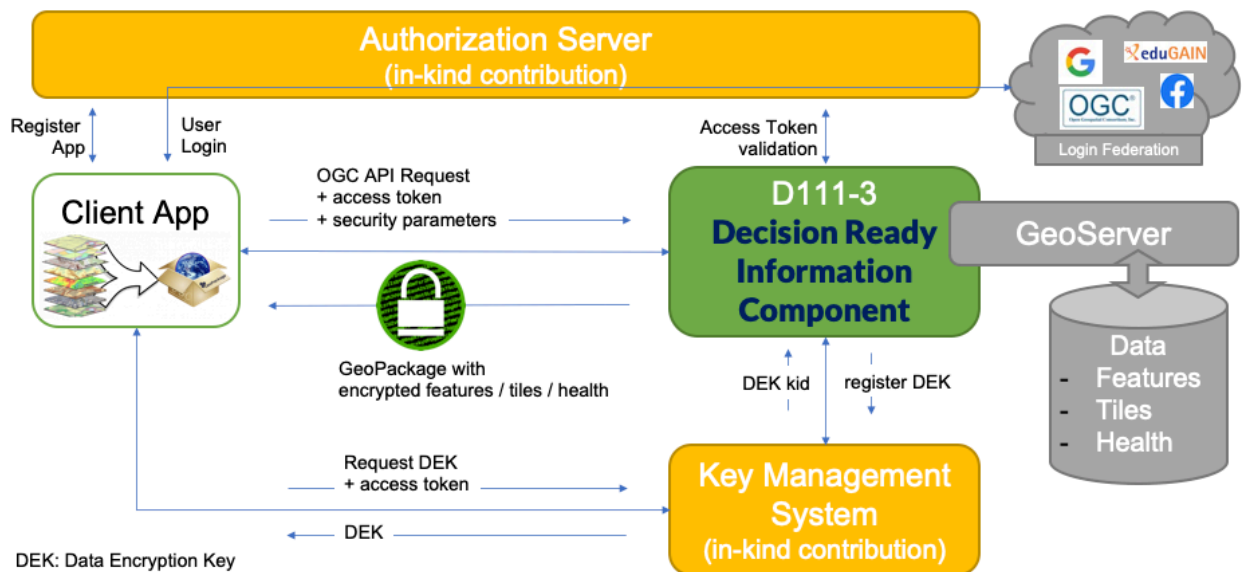
Figure 65 — D111-3 Component Overview and their messaging

5.10.2. Authorization Server

The Authorization Server used for the OGC Disaster Pilot '21 is an Authentication as a Service offering by Secure Dimensions GmbH that operates under the name of AUTHENIX. The service supports user logins from Facebook, Google, and OGC Portal as well as eduGain — the identity federation of Universities around the globe. In a nutshell, AUTHENIX brokers user identification, based on a crypto-identifier, available to registered applications via Bearer Access Tokens as defined in RFC 6749 (The OAuth 2.0 Authorization Framework). In addition, AUTHENIX makes personal information available to registered applications, based on different GDPR compliance policies and the user's consent, via the OpenID Connect protocol.

The major role of AUTHENIX in the Disaster Pilot is to allow the registration of trusted applications that are able to fetch (decryption) keys from the KMS. Via the key management portal of the KMS, applications are identified by the UUIDs received from AUTHENIX registration. This ensures to establish trust into applications.

The second role is to release Bearer Access Tokens to registered applications that allow authenticated users (login via Facebook, Google or OGC Portal) to request a GeoPackage with encrypted content and login to the KMS for managing access to their keys.

The third role is to verify Bearer Access Tokens for the GeoPackage Encryption Service and the Key Management System to ensure legitimate access.

The fourth role is optional but would allow applications to request personal information about the acting user, after the user has given consent. The ability of an application to obtain personal information also requires that the application links to a privacy statement that informs about the GDPR compliance.

AUTHENIX offers a rich API that covers automated OAuth2 compliant application registration, OAuth2 compliant access token verification and OpenID Connect compliant UserInfo requests.

Figure 66 — AUTHENIX API

Developers should follow the AUTHENIX OpenAPI page to explore API options.

NOTE For development and testing purposes, AUTHENIX Sandbox can be used.

5.10.3. DRI Component D111-3 (GeoPackage Encryption Service)

The Decision Ready Information Component D111-3 can be coined a Service that creates OGC GeoPackage with encrypted content upon request. D111-3, the GeoPackage Encryption Service (GES), is implemented as a GeoServer plugin that extends the OGC Web Map Service 1.1.0 and Web Feature Service 2.0.0 with Data Centric Security capabilities.

For the storage of encrypted content, the GeoPackage extension mechanism is used. Based upon the structure for Features and Tiles, the encrypted extension data structure is rather simple: one column in the data table(s) contains the encrypted content and the `gpkg_ext_keys` table contains the metadata of the symmetric key used for encryption or the protected key itself. Each row in the data table(s) uses a foreign key to reference the associated key. It is thereby possible to selectively encrypt individual features or tiles with different keys. This allows at a later stage to control access to partial information stored in one single GeoPackage. It is thereby possible that a GeoPackage contains information with different levels of sensitivity. It is also possible that individual tiles or features are encrypted with a different key, because of their geo-location (restricted area that requires different level of clearance). Annex Annex A introduces the GeoPackages Encryption Extensions for Features and Tiles in detail.

For demonstrating the capabilities of the GES, a simplified client application based on OpenAPI was deployed that can be used to create GeoPackage with encrypted content for OSM Places in the USA.

Figure 67 — D111-3 OpenAPI

Most of the request parameters are already fixed. The user must only provide a Bearer Access Token (by copy and paste) and a PIN or password (`key_challenge` parameter) that is required for modifying access conditions of the generated key via the KMS.

Figure 68 — D111-3 Requesting GeoPackage with encrypted Features

The user can provide a UUID identifier for an encryption key that was previously generated and registered with the KMS via the `key_id` parameter. If this options is used, the encryption is based on this key.

The use of the `key_challenge` (, `key_challene_method`) and `key_id` is mutually exclusive.

5.10.4. Key Management System

The Key Management System (KMS) is an integral part for the use of GeoPackage Encryption Extension. The KMS used for the Disaster Pilot '21 is an extension to the KMS developed for the OGC Testbed 16 initiative. The implementation honors aspects from the NIST 800-57

recommendation regarding key life cycle and access control. To clarify the management of `keys` NIST 800-57 differentiate between DEK and KEK: A Data Encryption Key (DEK) is a symmetric key that is used for encrypting data whereas the Key Encryption Key (KEK) is used to protect the DEK. The KMS implementation honors this differentiation.

Even though the Key Management System for the OGC Disaster Pilot '21 supports the management of KEK and DEK, only the capabilities to manage the DEKs are outlined here.

In order to keep the keys secure during storage, the KMS applies its own encryption to the key value before stored in a local database.

5.10.4.1. Data Encryption Key Creation

A key's lifetime begins with its creation. This can take place in an encryption service like the D111-3 service, offline or via the KMS admin interface. After the time of creation, when no data has being ciphered yet, the usage of the key is `encryption`. At the stage of encryption, the key might not be available for decryption purposes.

Figure 69 — KMS Input form for Creating a Data Encryption Key

When creating a DEK via the KMS, the user must provide different types of information, as below.

- `PIN` or secret that is required when adopting the access conditions of the key via the KMS Admin Portal.

- `Audience` is the UUID of the application that gets immediate access to the key (if the `Check to activate key` option is selected). Additional applications can be added via the KMS Admin Portal). In the case where the D111-3 service shall be used for creating encrypted content based on the created key, the GES' UUID must be provided. If AUTHENIX is used as the Authorization Server, the UUID of the application(s) can be found <u>here</u>.

- `Expires` determines the initial expiration time of the key. The expiration time can be changed via the KMS Admin Portal.

- `Chiper Algorithm` is a list of symmetric ciphers as illustrated in Figure 70. The top of the list is used with JWK and the list identified with URIs is used specifically for XML Encryption.

- `Check to activate key` option allows to make the key available immediately to the application specified above and until the expiration time.
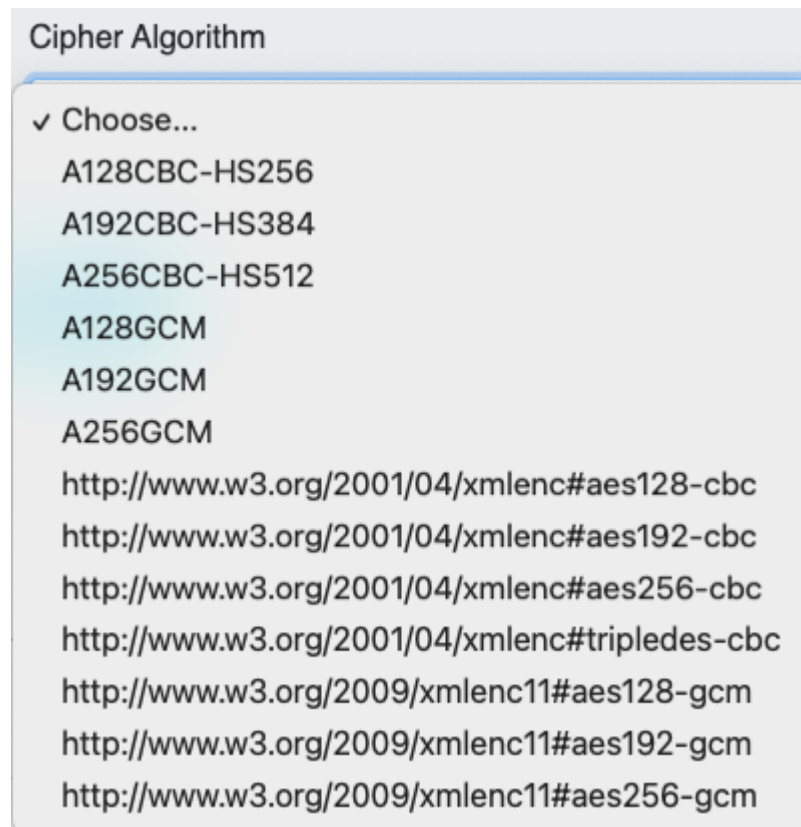
Figure 70 — KMS Ciphers for Data Encryption Key

For automatic DEK registration, the KMS offers an <u>API</u> that is described in OpenAPI.

Figure 71 — KMS API

For the automated registration of one DEK(s), the POST operation can be used. The response will contain the key identifier(s) that can be used with the GET operation to fetch the key details including the key's secret. The PUT operation supports the registration of an existing key where

the client depicts the key identifier beforehand. The key id (kid) must be a valid UUID. In case that another key with the UUID already exists, the response is HTTP Conflict. Compliant to the NIST 800-57 recommendation and to ensure that key identifiers are unique, the `DELETE` option does delete the key's secret and deactivates the key so that is can no longer be fetched via the KMS API. Even though the key could be reactivated via the Admin Portal, it is useless as the secret (`k`) value is set to `NULL`. However, no other new key can be created with the same UUID of any deleted key. This is important to prevent key forgery.

5.10.4.2. Data Encryption Key Registration

A data encryption key can be generated offline. Once data has been encrypted with the key and it comes to sharing the encrypted data, the key can be registered with the KMS. The KMS API supports different options:

- the Data Encryption Key (DEK) can be uploaded and the KMS created the key's identifier (`kid` in JWK terminology) via the DEK/addKey operation; and

- the user creates a UUID based identifier for the key and uploads key based on this identifier via the DEK/addKEyById operation.

5.10.4.3. Data Encryption Key Access Management

Once a DEK is registered with the KMS, the owner of the key can modify the access conditions after login. Once logged in, the KMS displays a list of all keys accessible to the user (see Figure 72).
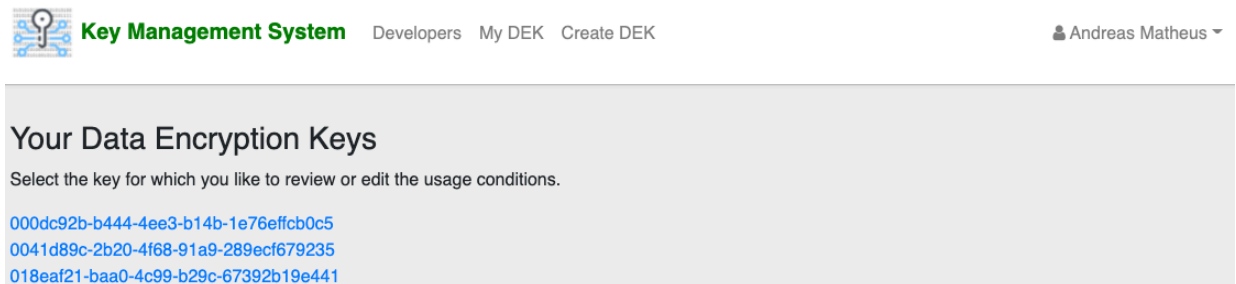
Figure 72 — KMS Showing list of "my" keys

The user can open the access management page by following the link for a particular key. As illustrated in Figure 73, the key owner has different options to constraint access.

Figure 73 — KMS Access Management Page

In order to change access conditions for the key, the user must provide the `PIN`. The blue `?` can be clicked to check if the PIN is correct. If so, the icon will change to a green check-mark; if incorrect, the icon will change to a red X.

The `activate` option controls whether the key can be accessed at all. If unset, the key cannot be fetched from the KMS, regardless of the conditions specified. This is a kind of emergency override.

The next option allows to make the key available to users based on their email address(es). Please note that the email address of a user depends on the login provider chosen when logging in via AUTHENIX. If the `OGC IdP` is used for login, the email address is equivalent to the email address shown on the "my" page of the OGC Portal; if logged in via Facebook or Google, the user's corresponding address must be used for sharing.

In case that the key shall be shared with user(s) but the users do not want to provide their email address(es), the use can provide their UUID. The UUID is displayed once logged into AUTHENIX. The user's identifier UUID is displayed at the bottom of the page. It is the value for the `Sub` attribute name. This honors the privacy of the users as coined in GDPR.

Any application that needs to obtain a DEK from the KMS must first be registered with AUTHENIX and the `client_id` UUID from the registration result must be put into the `application UUID` box. Any application that is currently registered with AUTHENIX can be found following the Operators link. For example, the OGC Token App used for various demonstrations has UUID `019b7173-a9ed-7d9a-70d3-9502ad7c0575` and the Compusult Mobile App developed for OGC Testbed 17 that supports decryption of GeoPackages with Encryption Extension (as defined in Testbed 17) has UUID `8d90bc42-4401-5f2a-9054-cb407a876ad8`.

Next, the user can specify the time window of access. The `from` time can be set to the future which is an important feature if encrypted data is packaged on some mobile devices and they are shipped to their final destination. While the devices are in transit, the associated decryption key cannot be fetched from the KMS.

In addition to the temporal conditions, also geospatial conditions can be added. The selection of multiple polygons allows to constraint access to keys for particular areas. This ensures for example that mobile devices can only obtain the decryption key if within the 'green' area. If this geospatial access conditions is enabled, the request to getKeyById operation must include the `subject_location` parameter.

5.10.5. The OGC Token App

As illustrated in Figure 65, all different services co-play based on Bearer Access Tokens created and validated by AUTHENIX.

For supporting development and demonstration of different aspects of the D111-3 component, the OGC Token App was developed during OGC Testbed initiatives 15 and 16. It is a simple One Page Web Application that allows to display a Bearer Access Token to be used via copy & paste.

Figure 74 — OGC Token App

The login via the `OGC IdP` enables access tokens to refer to user's profile information stored in the OGC Portal. To illustrate the OpenID Connect claims for an access token, AUTHENIX's OpenID Connect UserInfo endpoint can be used (see Figure 66).

As a result of the UserInfo request (e.g., requesting only scope `OGC`), the `ogc-is-member-of` claim reflects the active project memberships within OGC as illustrated in the example below.

```
{
  "sub": "95d3827d-efef-3052-a2df-46d1c7ce77a5",
  "ogc-is-member-of": [
    "Architecture DWG",
    "Security DWG",
    "OGC Architecture Board - OAB",
    "OGC API - Common SWG",
    "GML 3.3 SWG",
    "Sensor Planning Service (SPS) 2.0 SWG",
    "Sensor Model Language (SensorML) 2.0 SWG",
    "Sensor Observation Service (SOS) 2.0 SWG",
    "O&M SWG",
    "OGC API - Processes SWG",
    "GeoXACML SWG",
    "CityGML SWG",
    "SWE Umbrella SWG",
    "GML Umbrella SWG",
    "OWS Context SWG",
    "Features API SWG",
    "SensorThings SWG",
    "Land and Infrastructure SWG",
    "OWS Common - Security SWG",
    "Blockchain and Distributed Ledger Technologies DWG",
    "IP Pool",
    "OGC Testbed-17",
    "Disaster Pilot 2021",
    "OGC Web API Guidelines sub group"
  ]
}
```

Figure 75

If the user is a member of this Disaster Pilot Project, they find the entry `Disaster Pilot 2021` in their list.

5.10.6. Establishing Access Control to Services

For the OGC Disaster Pilot, access can be constrained via AUTHENIX Bearer Access Tokens and the `ogc-is-member-of` claim. Only if the user logs in via `OGC IdP` and has subscribed to this project via the OGC Portal, the membership to `Disaster Pilot 2021` is present. Therefore, access conditions for this pilot can easily be setup based on the `ogc-is-member-of` claim. A proof can be easily conducted if a user logs in via Facebook or Google. Then, this claim does not exist.

For the D111-3 component, access can easily be constrained to this particular condition. It is thereby ensured that GeoPackage(s) with encrytped data for the Disaster Pilot can only be generated by users that are member of the `Disaster Pilot 2021` initiative.

5.11. Analytical Processing Component (Skymantics)

5.11.1. Background

The main goal of Skymantics' participation in the Disasters Pilot 2021 was to analyze the impact of road and transport networks in the management of disasters. On one hand, a proper management of roads during disasters hazards can save lives: according to the Centers for Disease Control and Prevention, over half of all flood-related drownings occur when a vehicle is driven into hazardous flood water (https://www.weather.gov/safety/flood-turn-around-dont-drown). On the other hand, transportation has a direct impact on the prompt distribution of food and medical supplies, as well as emergency and relief teams, helping to save lives during disasters hazards.

In order to test the different scenarios, Skymantics would deploy a Road Risk Estimator to consume flooding information produced by D101, as well as other data sources, and to transform them in real time into an estimation of risk for all roads in the area, using Skymantics governance automation system. This system would then decide which roads should be closed, which should be kept open only for emergencies and which could remain open. In addition, Skymantics would deploy a routing engine for the different scenarios to calculate optimal routes and driving times, as well as a routing client to visualize the generated routes, based on the client developed in OGC Testbed 17 (D175).

Skymantics started a close cooperation with RSS-Hydro and Safe Software to consume their ARD outputs, as well as with HSR.health to define new DRIs to produce and their applicability in disaster scenarios.

5.11.2. Decision Ready Indicators

As an initial step, a research was carried out to find the most relevant decision ready indicators (DRI) pertaining road networks and transport management and to document the recipes that would generate these DRIs. The three most important recipes documented this way were "Manage flooded roads", "Update supply routes," and "Estimate distance to care." These three recipes were implemented and tested in three different scenarios.

Table 7

RECIPE	SCENARIO	INPUT ARD	PROCESS	OUTPUT DRI	TYPICAL THRESHOLDS	SUGGESTED ACTIONS	EXPLANATIO
Manage flooded roads	Flood	Roads dataset; Extension & depth of current flooding; Terrain topography (road elevation); River and flood gauges (water level); Inputs from public and first responders (water level)	Find roads affected by flooding	a) Road X flooded but passable; b) Road X flooded and only passable by special vehicles; c) Road X flooded and impassable	a) <150mm flood depth; b) 150-300mm flood depth; c) >300mm flood depth	a) Reduce estimated road speed to 70-20 kmh; b) Reduce estimated road speed to 20-2 kmh. Block road for public. Notify traffic apps; c) block road for all transit	With 150 mm depth, water can enter a small car's exhaust pipe. With 300 mm depth, cars can start floating. Several articles point to a substantial traffic speed reduction based on depth.
Update supply routes	Health/ Flood	Updated roads dataset generated in "Manage flooded roads" recipe; Location of supply depots and supply destinations; Traffic status	Recalculate supply route	a) Faster alternative route; b) Supply route impassable; c) Supply route jammed	a) Alternative time < Updated route time; b) Impassable; c) Jammed	Calculate new route. Notify of new supply route.	Supply lines need to be open at all times.

RECIPE	SCENARIO	INPUT ARD	PROCESS	OUTPUT DRI	TYPICAL THRESHOLDS	SUGGESTED ACTIONS	EXPLANATIO
Estimate distance to care	Health/ Flood	Updated roads dataset generated in "Manage flooded roads" recipe; Population density and geometries of the Administrative Regions; Location of hospitals	Calculate distance to closest hospital for each administrative region	a) Distance to care poses a low risk; b) Distance to care poses a medium risk; c) Distance to care poses a high risk	a) Less than 5 minutes; b) Between 5 and 30 minutes; c) More than 30 minutes	Evacuate most vulnerable population. Set up temporary hospitals close to high risk districts.	Mortality rises if distance to care is over 5 minutes. It rises even more if distance to care is over 30 minutes.

It is important to note that the output DRI of the first recipe becomes one of the input ARD for the other two recipes.

5.11.3. Technical architecture

The core of the Skymantics component is its routing engine. It downloads and stores the road network from OpenStreetMap. The routing engine is based on the open source software pgRouting, that is fully compatible with OpenStreetMap datasets. Skymantics routing API connects to the routing engine and offers the possibility to calculate routes following the specifications of OGC API Routes. Using Skymantics routing client (open source, developed in OGC Testbed 17 as D175), a user can easily request new routes or load pre-calculated routes using a graphical interface in a web browser.
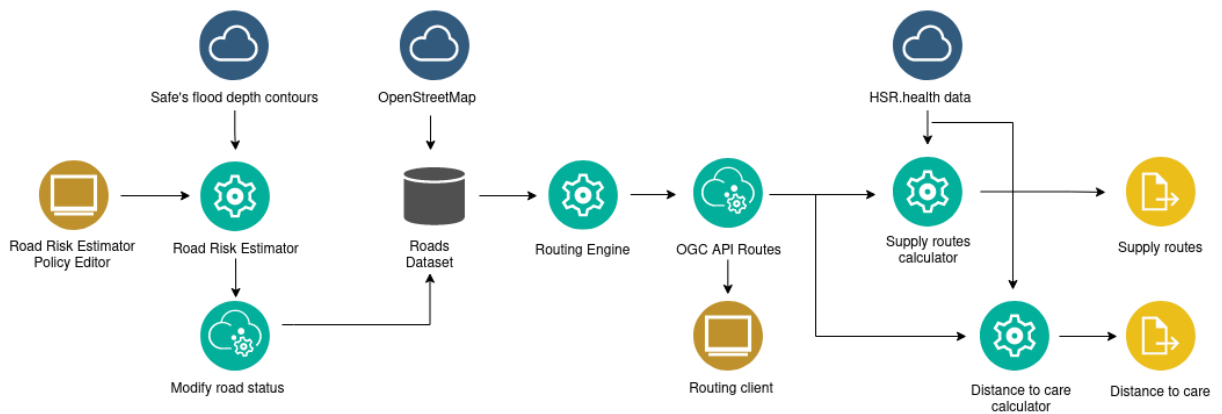
Figure 76 — Technical Architecture Diagram: Technical workflow of Road Risk Estimator

This core is extended in two ways to adapt to the needs of the Disaster Pilot 21:

a) A Road Risk Estimator is added to process flooding data and modify the status of roads; and b) Combining the routing capabilities with health data, new applications are develop to generate and update supply routes as well as to calculate distance to care indicators.

5.11.3.1. Road Risk Estimator

Safe Software's FME component generates a set of flood depth contours with up-to-date information of a flooding and writes it to a Geopackage. Skymantics' Road Risk Estimator receives this data and pre-selects the roads that are being affected by the flooding. It applies the values for thresholds to decide at which flood level a road should be open only for emergency vehicles or closed for all transit.

The Road Risk Estimator will use these thresholds and the current flood level at each road to decide the status of each road, as well as the estimated traffic speed in case it remains open. Using a REST controller the Road Risk Estimator sends the updated road data to the PostgreSQL DB where the main road data is located. The pgRouting routing engine will use this updated data to generate new routes, avoiding closed roads.

A user can modify the Road Risk Estimator parameters using the Policy Editor, a web application that allows editing the scenario's thresholds such as flood depth, distance, etc. The Policy Editor allows for communication between the frontend and the Road Risk Estimator through the usage of REST services and serlvet technology. Updating the new parameters will reprocess the road segments with the new thresholds. The new road dataset will be uploaded to the database and will be used by the routing engine in future requests.

5.11.3.2. Health related applications

Two new applications were developed that made use of the routing engine, through its OGC Routing API, to calculate supply routes from a depot to hospitals and to calculate distance to care from districts or regions to the closest hospital. The health data required, such as hospital locations, supply depot locations, or district boundaries, are downloaded from HSR.health GeoNode, The resulting processed data are delivered in GeoJSON format for further analysis by health experts.

5.11.4. Scenarios and implementations

5.11.4.1. Red River flooding; road management

The Red River is a large basin in central North America that spans the US/Canadian border. The river flows from the Dakotas and Minnesota to Manitoba and Hudson's Bay draining almost 300,000 sq kms. The Red is susceptible to flooding due to its flat slope and typically large snow melts & ice jams in spring.

Based on Earth Observation monitoring of the historical flooding in the Red River basin, RSS-Hydro developed a flood time series model. Safe Software's FME component converted this model into Integration Ready Data, consisting on a set of flood depth contours written to Geopackage and GeoJSON for use by downstream applications and published to the DP21 GeoNode.
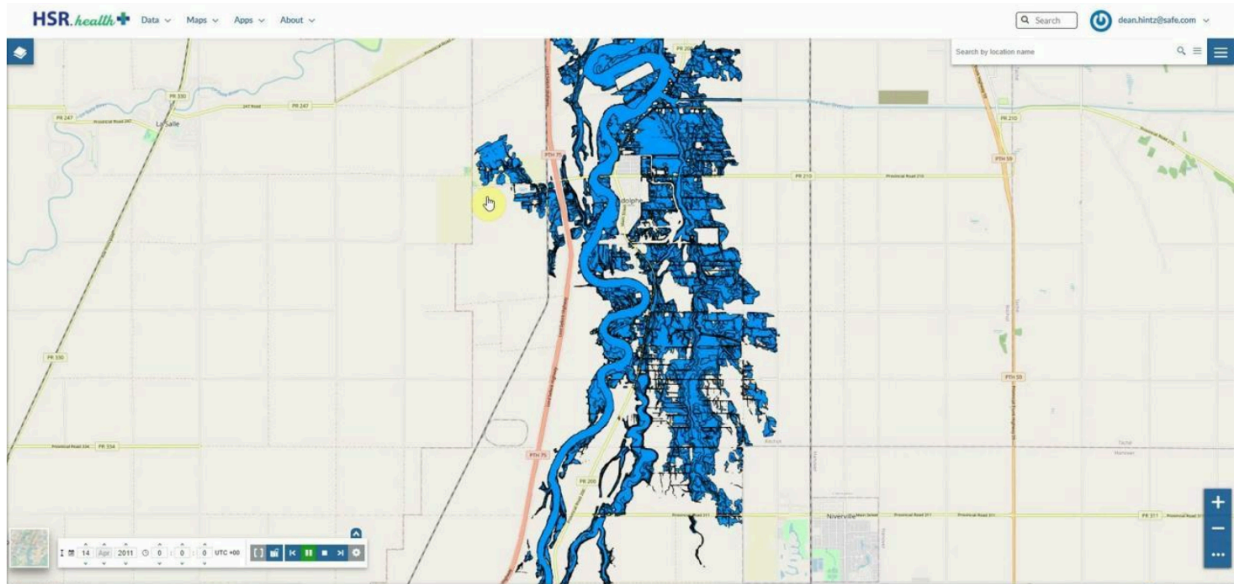
Figure 77 — Red river: flood depth contours generated by Safe

Skymantics deployed a routing engine with OpenStreetMap data for the river basin, which in combination with its routing client provided a full environment to calculate routes in the area. The following image shows the optimal route from Osborne to Ste. Anne in a normal scenario, without flooding.
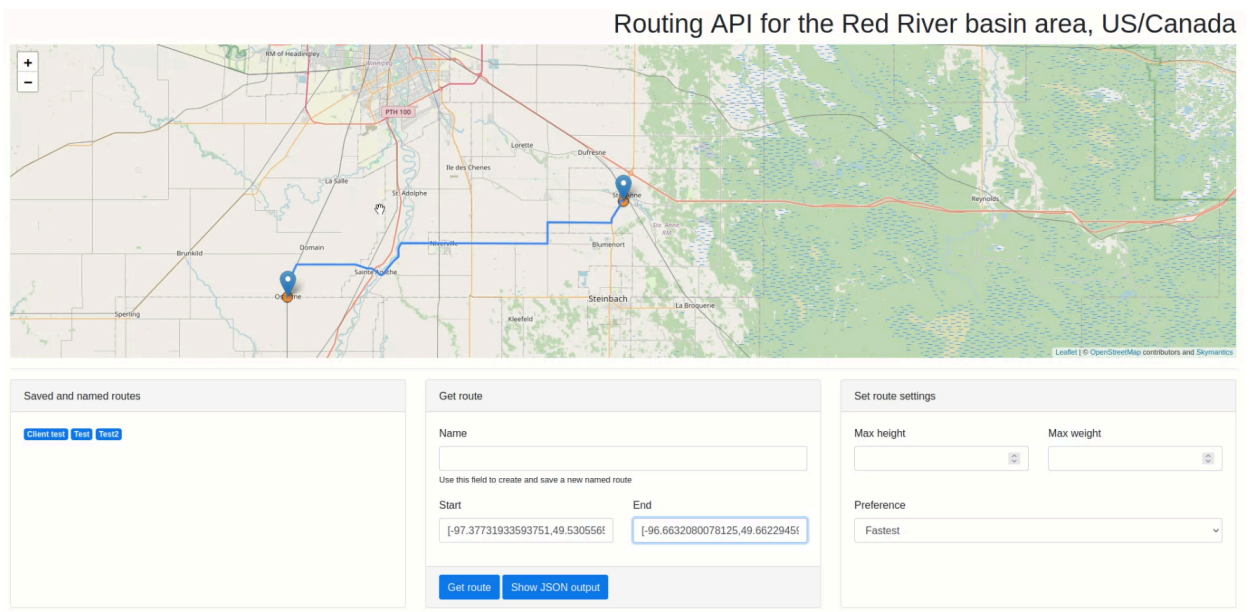
Figure 78 — Red river: optimal route without flooding

Prior to a flooding, a decision maker can select the security thresholds that will control road closing in the routing engine, both for public and emergency vehicles, adapting the behavior to the local context. These thresholds can be modified graphically using the Road Risk estimator policy editor.
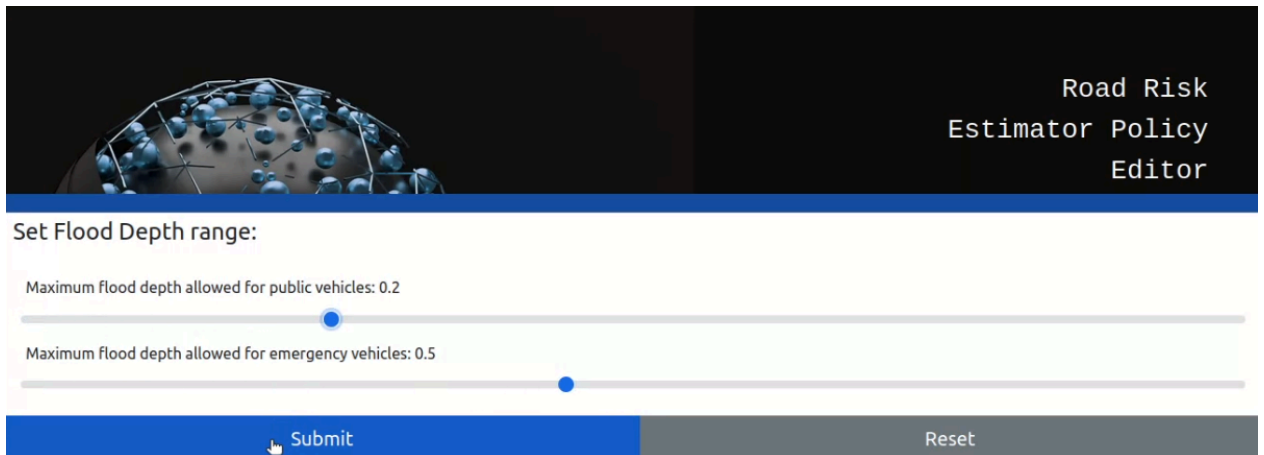
Figure 79 — Red river: Road Risk estimator policy editor

When a flooding occurs, the Road Risk estimator receives the contours and depth of the flooding at each moment and it processes this information to asses which roads are affected and the depth on each affected road. If the flood depth surpasses a security threshold, defined by the Road Risk estimator policy editor, then the road will be considered blocked and new routes will be generated avoiding it. The following image shows the route from Osborne to Ste. Anne when the flooding is close to its peak.
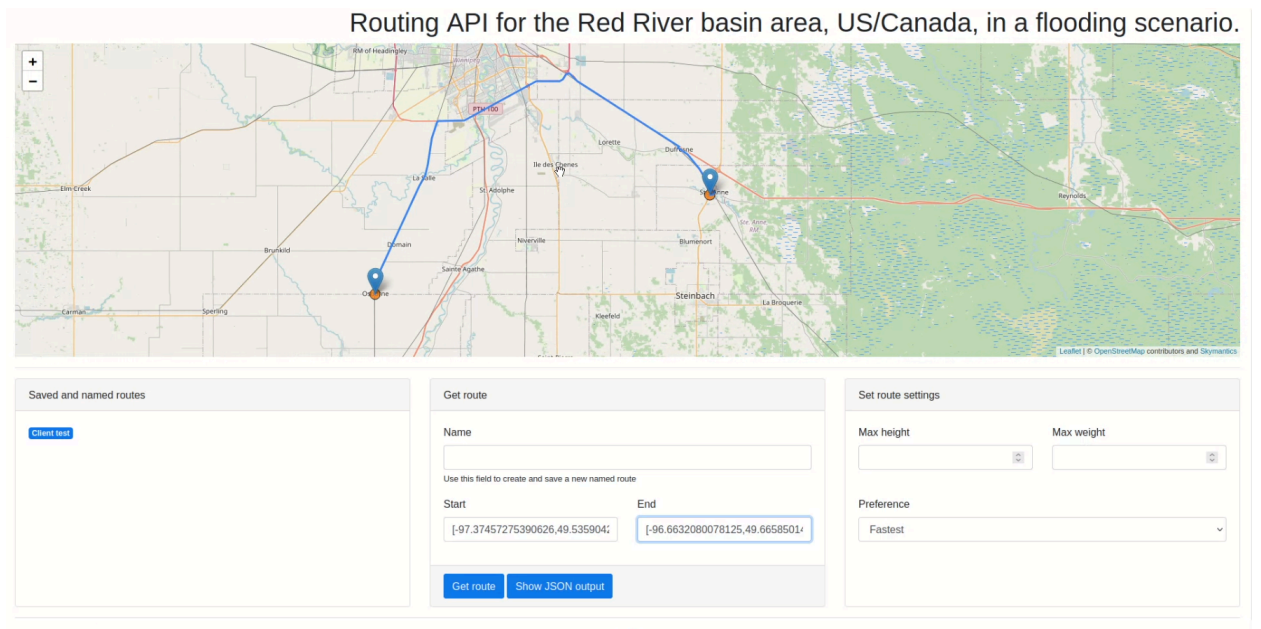
Figure 80 — Red river: route with flooding

5.11.4.2. Rimac river basin flooding; supply routes

In the Peru scenario, the implementation combined HSR.health's Medical Supply Needs Index with Skymantics' capability to generate supply routes. The Medical Supply Needs Index is used to determine when and how much supply needs to be delivered to each hospital and clinic. In case of a natural disaster, the medical supply needs calculation can be adjusted based upon increased volume of patients (due to the disaster) and based upon a shift in load on the medical systems if a hospital or clinic becomes inaccessible due to the natural disaster.

In this scenario, a local railway station acts as the main supply depot for the local hospitals. The following image shows the routes (purple lines) used to directly supply five of these hospitals in a normal situation, without any flooding occurring. Two hospitals (marked with a red circle) will have their supply routes disrupted during the flooding.
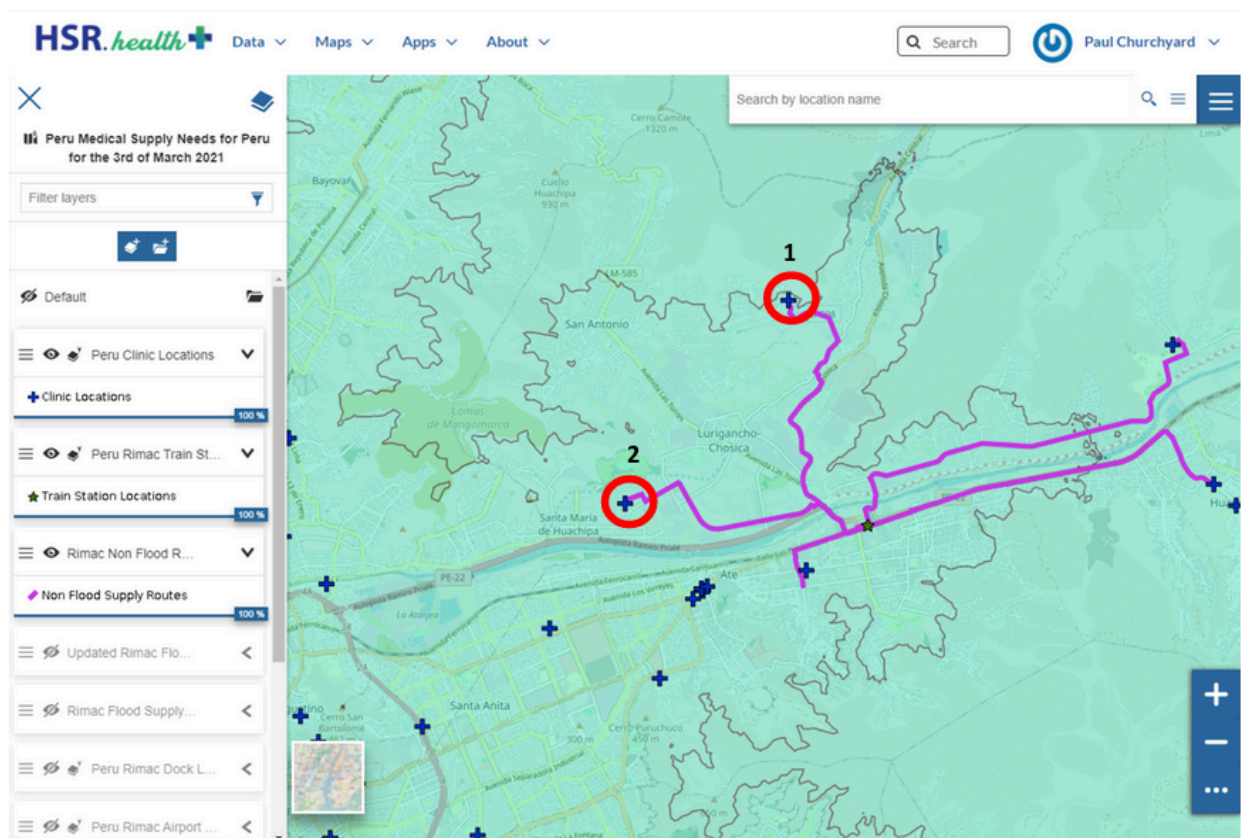
Figure 81 — Rimac river: preflood medical supply routes

When the flooding occurs (its extent displayed in blue), several supply routes are affected. One can easily be re-routed with minimum impact, but the routes supplying hospitals 1 and 2 need very long detours to keep them operational. In this case, a better solution is to direct their patients to nearby hospitals 3 and 4, which can be re-supplied easily with relatively short routes from the railway station.
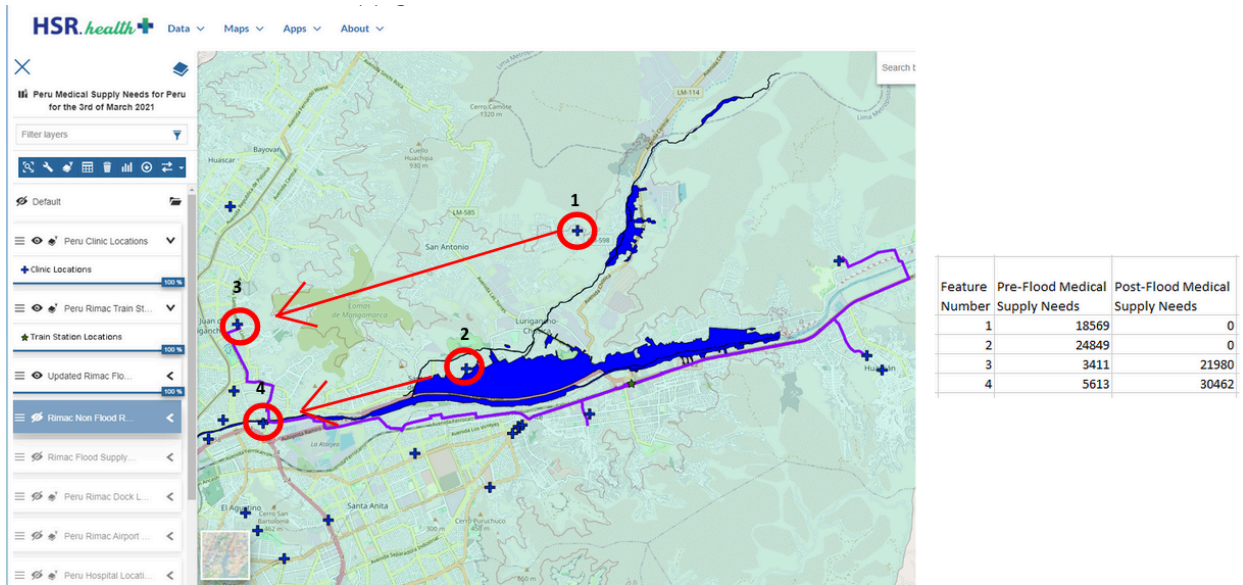
Figure 82 — Rimac river: recalculation of medical supply routes due to flooding

The Medical Supply Needs Index is also recalculated for all four affected hospitals, keeping their values up-to-date.

5.11.4.3. New Orleans pandemic; distance to care

The third scenario focused on a pandemic and health use case in the United States. New Orleans, Louisiana was chosen as the area of study as Louisiana has census tract level of pandemic data and is an important use case to build off of in future pilots due to the frequency of coincident natural disasters that happen there.

The analysis in this scenario centered on experimenting with a new indicator, distance to care, defined as the distance to the nearest available hospital. According to experiences in other parts of the world, a distance to care below 5 minutes is optimal, whereas distances below 30 minutes pose a moderate risk, and distances longer than that pose a serious risk for the population.

In collaboration with HSR.health, a pandemic scenario was simulated, where about one third of the hospitals in the city were saturated and could not attend to new patients. For every district the routing engine calculated the average distance to the nearest available hospital. The results showed no area in high risk, although several were in moderate risk, depicted in yellow.
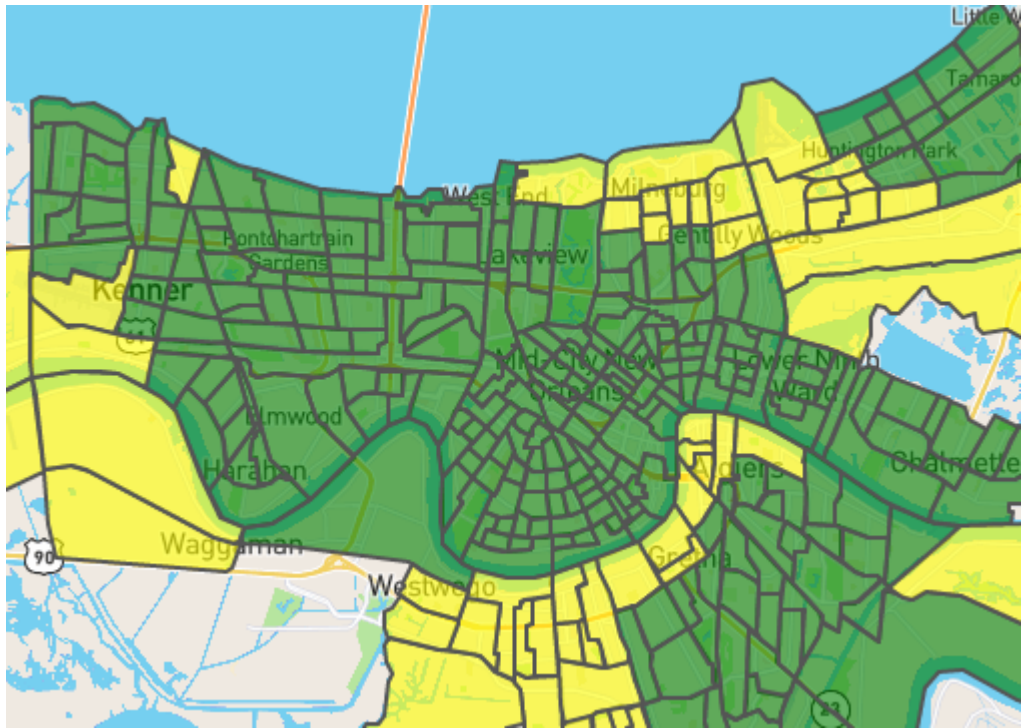
Figure 83 — New Orleans: distance to care in a pandemic scenario

Such a result can help prioritize the saturated hospitals that should be relieved, or in which areas new temporary hospitals could be deployed.

5.11.5. Challenges and Lessons Learned

- The distinction between ARD and DRI is not absolute: DRIs produced by one recipe can be the ARD inputs for another.

- In order to estimate the roads affected by a flooding at a specific moment, apart from the flood extent, the flood depth is a critical factor. However, in some scenarios it might be difficult to gain access to satellite imagery of sufficient accuracy, which was the case in the Rimac river scenario. Other sources of data can be considered to estimate the flood depth as an alternative or a complement to satellite imagery, such as river and flood gauges, inputs from public and first responders (such as the one proposed by GISMO), or modelings based on terrain topography.

- As demonstrated in this pilot, transforming flooding ARD (flood contours and depth) into DRI (roads that should be closed) needs a series of parameters. The value of these parameters depend on the local context and should be adaptable. For example, the maximum flood depth for emergency vehicles will be higher in municipalities that have invested in equipment that is capable to wade deep waters. The maximum flood depth for public vehicles, which depends on the typical ground clearance of cars, will be higher in rural America than in urban Europe.

- The case of flooding disaster prevention in roads was considered, particularly for the scenario of flash flooding in roads without adequate drainage that can pose a mortal risk

for drivers. However, it does not seem to be feasible using satellite imagery and it would require visual inspection on the field by expert personnel.

- The integration among different components in the data flow has been tested at the level of data format, but tests were not complete at the level of live data, APIs, and processes. In the case of flood contours processing and road closure management, Skymantics' Road Risk Estimator could complete the task for the Red River scenario in a matter of seconds, estimated between 10 and 30 seconds depending on the flooding moment. This performance is considered acceptable based on the low frequency of ARD updates and proves the technical feasibility of the concept. In any case, more integration and testing would be necessary at the service level to fully demonstrate the real-time flood information.

- The potential to use routing and travel time calculations for disasters was tested successfully in two health related applications, one for supply routes and the other to estimate distance to care for entire districts. There is still a lot of potential to be uncovered combining routing and distance calculation capabilities in specific domain applications, particularly in planning and relief. For example, finding the optimal location for shelters, supply depots or temporary hospitals based on minimal travel distance for a targeted vulnerable population.

- The case of landslides poses a challenge to the prediction and terrain analysis tools. However, from the point of view of road management, it was considered a simplification of the flooding scenario, solved by just blocking the roads affected after the hazard. The case of wildfires can be more complex and of great interest for future work, with several different ARDs to be considered for a proper road management, starting with roads with visibility affected by smoke, to the decision of closing a road based on the distance to fire and the expected evolution of the wildfire.

5.12. Communications Mobile Awareness | Web/Desktop Visualization Component (StormCenter)

5.12.1. Introduction

StormCenter Communications (StormCenter) has contributed to the Disaster Pilot by initiating an instance of GeoCollaborate, StormCenter's real-time data access, sharing and collaboration platform to test data access from DP21 participants that have generated data layers to be delivered into decision making environments. GeoCollaborate was demonstrated numerous times and put to use use by pilot project participants to demonstrate how this collaborative environment can fuel trusted data use and data-driven decision making. GeoCollaborate provides a collaborative environment that includes portability and scalability for testing in remote geographic locations such as Peru, Manitoba Canada, and anywhere in the U.S. or around the world. StormCenter addressed components D114-5 and D116-7 to deliver decision-

ready information (DRI) directly to mobile devices and decision maker's desktop computers to improve situational awareness and data-driven decision making.

The platform has played a critical role of delivering ARD and DRI into devices and computers to place all participants on the same map at the same time in a collaborative environment. The image below is an example of a Leader and Follower using disparate web browsers to demonstrate how situational awareness and decision making can improve when all participants are on the same map at the same time.
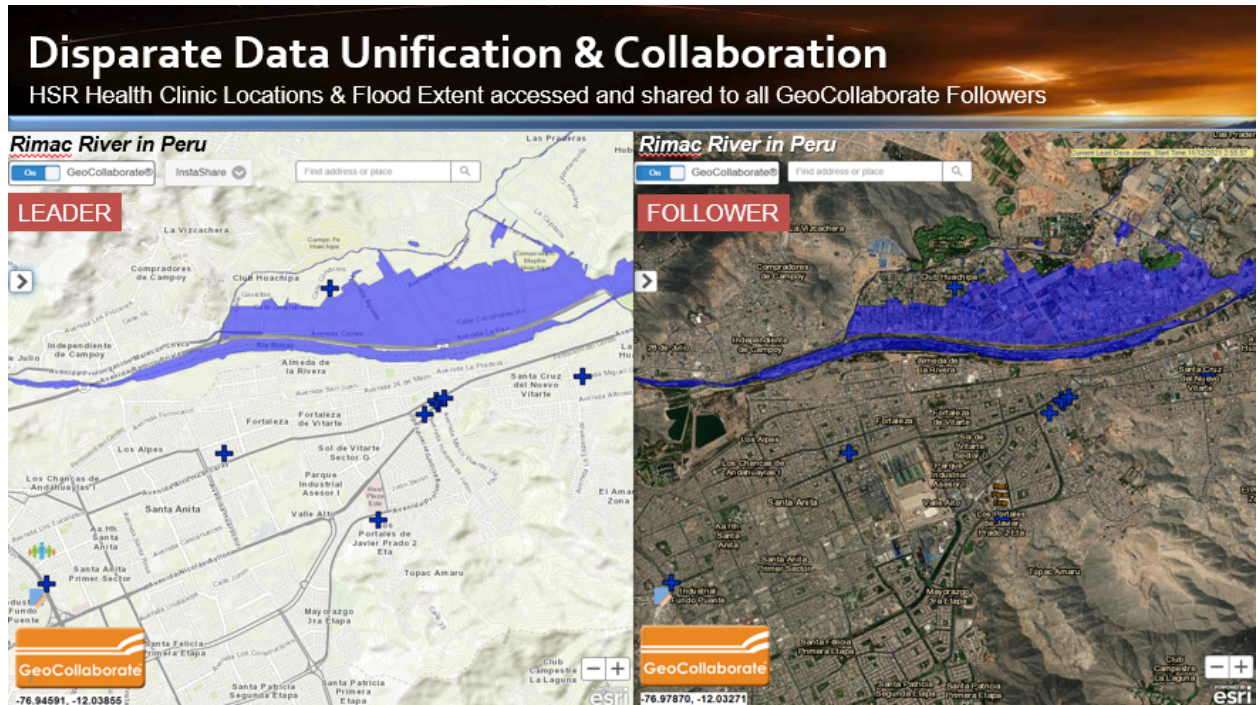
Figure 84 — The authorized GeoCollaborate Leader is sharing data with the Followers in a real-time collaboration session.

- GeoCollaborate, a collaboration-as-a-service (CaaS) technology, achieves total and true commonality and permits collaboration across all stakeholders accessing trusted decision-ready information (DRI) using interactive web maps, GIS platforms, data portals, hubs, or common operating pictures (COPs). OGC data compliance ensures cross platform interoperability as GEoCollaborate operates on any device and across platforms.

- Supporting open data standards to ingest, process, store, disseminate, and visualize raw data, ARD, and DRI from fellow participants including Safe Software, RSS Hydro, and Skymantic, as well as the collection and ingestion of on-the-ground data from field personnel, including but not limited to: health data, location and availability of existing healthcare facilities and resources, transportation information, road conditions, location of temporary or field health facilities and resources, environmental conditions, and other data relevant to situational awareness.

- In the same way that map services bring real-time map information to maps, GeoCollaborate is a network service that permits real-time data sharing and collaboration across an unlimited number of disparate web maps and platforms. It allows anyone to securely author the content of a lead web map, share content, and collaborate in real

time or offline with other follower web maps with nothing more than a browser and a network connection. At the end of each collaboration session, the full set of original datasets visualized, and annotations created during the collaboration session remain active on each collaborator's device until turned off by the leader, allowing session participants to perform further analyses and deliver other presentations. GeoCollaborate's approach to this task has been to access published structured geospatial data using the JSON-LD, KML, WMS formats through a GeoServer OGC API implementation. The JSON-LD context was defined based on Schema.org vocabulary.

For many years, decision makers in the field have desired access to data that does not require downloading large image files due to limited bandwidth. GeoCollaborate delivers a way to interact with decision makers, data analysts and data producers in a collaborative environment so data can be processed and produced as Analysis Ready Data (ARD) or Decision Ready Indicators (DRI) AND delivered into the field or on-site operations center using low bandwidth connections.

5.12.2. Description of the contributed component(s) and role in the Pilot architecture

The figure below shows where GeoCollaborate fits into the DP21 Pilot. The schematic shows the red circles where GeoCollaborate accesses and delivers OGC web service data formatted ARD and DRI into platforms and devices to advance situational awareness and decision making. Operating in the cloud, GeoCollaborate accelerates data access and transfer for immediate connectivity between the data provider and the data user.
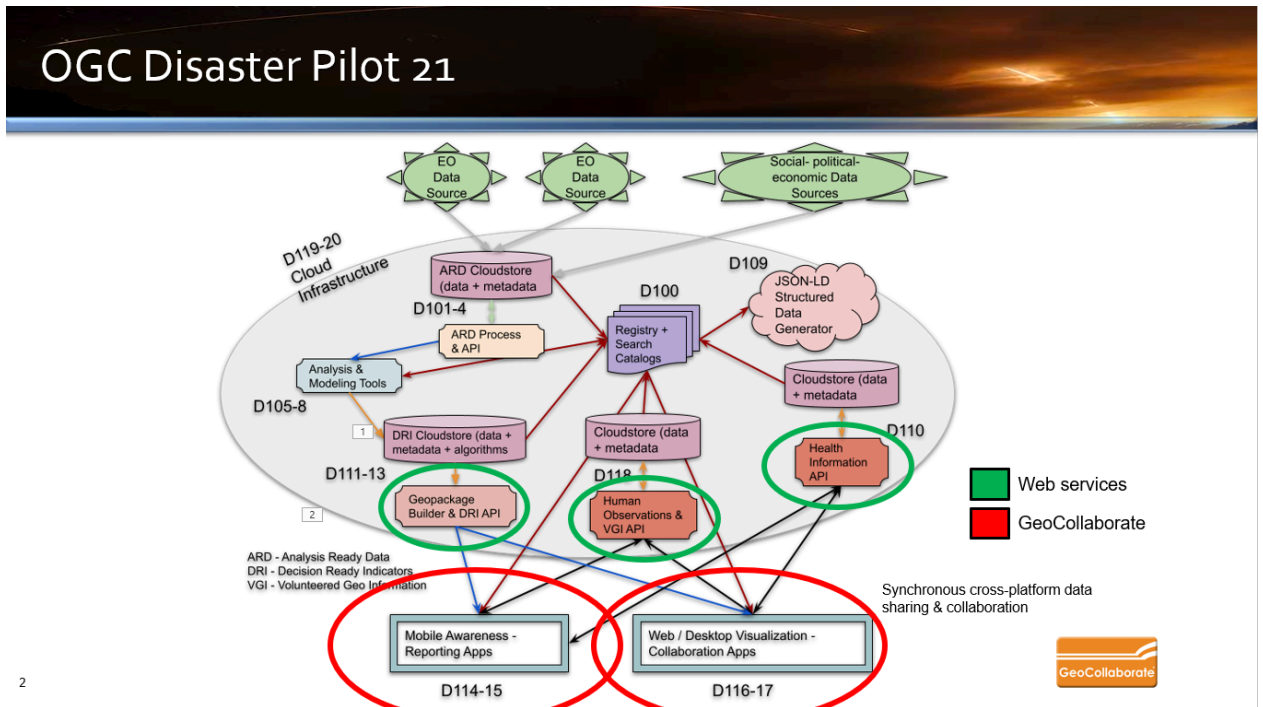
Figure 85 — GeoCollaborate addressed D114-117 which focused on delivering data services into mobile and desktop visualization platforms while engaging

them in real-time cross-platform synchronous collaboration sessions. This was demonstrated successfully many times during OGC stakeholder meetings.

By focusing on the Web/desktop client application that can interact with cloud-based server components to discover, request, and download both ARD and DRI products for analysis and visualization, GeoCollaborate can directly support use cases and specific challenges identified by the sponsors of this pilot project should the desire to perform exercises arise. GeoCollaborate supports virtual collaboration between distributed stakeholders both in the interpretation of such datasets and in the design and selection of "recipes" for the generation of decision-ready information products. GeoCollaborate can function as a "virtual fusion center" where collaboration sessions are mainly synchronous, but also provide asynchronous interaction.

The figure below shows how GeoCollaborate can access web services from the HSR.Health Geoplatform and share them across any device. Cloud-based data services are very efficient in supporting a large number of collaborators on any device as they can scale based on the number of participants. We describe the use of OGC data standards as 'turbo-charging' data interoperability and collaboration.

Figure 86 — GeoCollaborate accessing HSR.Health GeoPlatform
web service and sharing it across all follower's web maps.

GeoCollaborate also can take advantage of APIs that provide interactive geospatial queries such as population. StormCenter Communications worked closely with the NASA Socioeconomic Data Applications Center (SEDAC) at Columbia University to enable real-time extraction of population estimates from anywhere in the world based on the Population Estimation Service (PES) which is a Web-based service for estimating population totals, basic demographic characteristics, and related statistics within a user-defined region. It enables users of a wide variety of map clients and tools to quickly obtain estimates of the number of people residing in specific areas without having to download and analyze large amounts of spatial data.

The image below is an example of the Leader of a collaboration session making a population extraction query by drawing a freehand polygon. GeoCollaborate then communicates with the SEDAC API to request the population in the drawn polygon and display it on the Leader's instance and share it across all Follower instances in real-time. Also displayed is a graph depicting population changes over a period of time based on census data. These charts can be closed and re-displayed by clicking on the green free-hand polygon. Pictures below is an example of determining how many people may be impacted by the flood inundation data layer produced by HSR.Health.
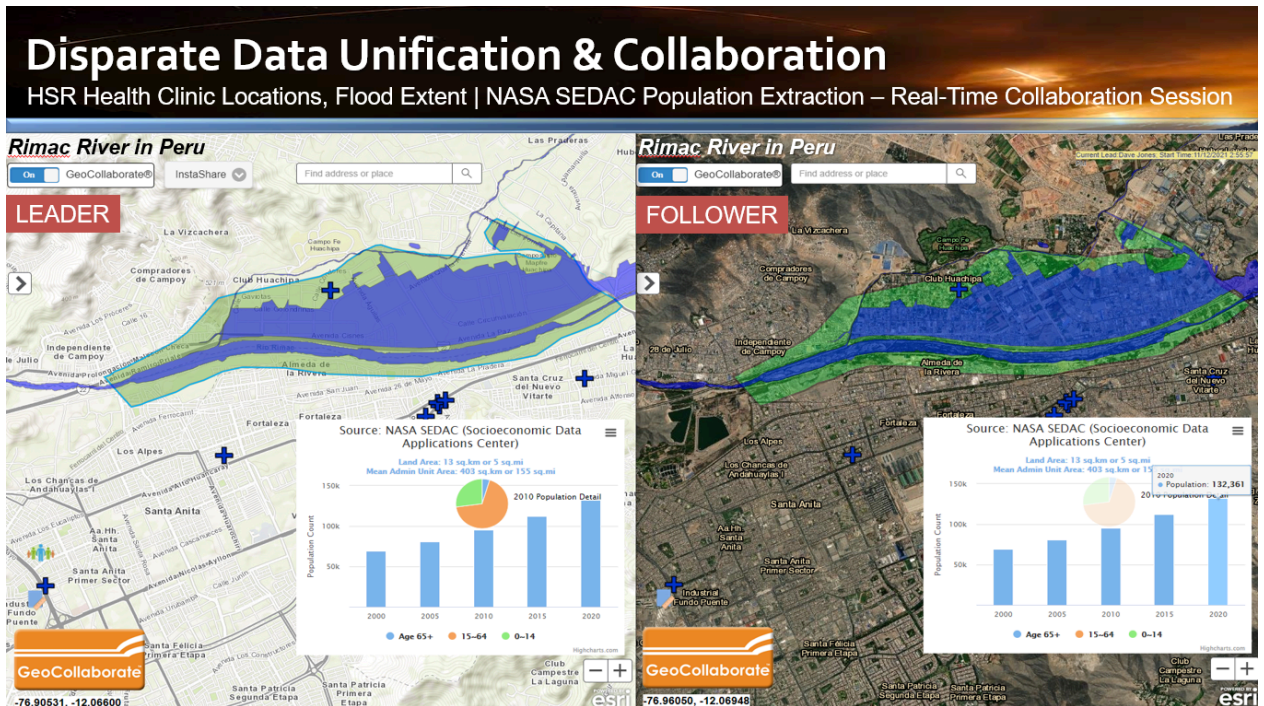
Figure 87 — GeoCollaborate accessing NASA SEDAC population data as a web service through their API and sharing it across all follower's web maps. Once a population query is requested by the Leader of the GeoCollaborate session, the results are then shared in real-time across all 'Follower' instances and with all dashboard users.

5.12.3. Technical design and implementation

The DP21 identified a challenge that data can not easily be delivered into the hands of decision makers in the field and that bandwidth is a limitation for delivering high data volumes into disaster areas. This is accurate and was a driving requirement when GeoCollaborate was developed under the US Federal Government's Small Business Innovation Research (SBIR) Program. StormCenter Communications answered a grand challenge from NASA that required a technology that had not at the time existed, to help federal agencies access and share geospatial data across platforms. This multi-year effort led to the creation of GeoCollaborate and this development effort has been significant enough to result in GeoCollaborate reaching Phase III status. As a result of the SBIR contract, GeoCollaborate is a patented technology and is able to be implemented across any US Federal agency to accelerate the accessing and sharing of geospatial and other data across platforms providing data producers, curators and providers with

the ability to deliver data as a service while establishing cross-platform interoperability to speed situational awareness and decision making.

GeoCollaborate can now be implemented across any platform rapidly to establish a data sharing and collaboration environment that can provide ARD and DRI into environments where decision makers can act on trusted data faster than before.

5.12.4. Scenario workflows and recipes

The workflows and recipes developed under DP21 produced data layers that have value to improve situational awareness and decision making while introducing indicators that could be valuable in monitoring hazard evolution. GeoCollaborate delivers upon its DP21 proposal by accessing and sharing those data layers across multiple platforms using a web interface. Anyone who has access to an internet connection (via any means) and a web browser, can connect to a collaboration session and participate in a data sharing session. These sessions can be led by subject matter experts (SME) who have knowledge of the specific data layers being shared. These data layers can include satellite imagery, sensor point data, crowd-sourced observations, model output, Artificial Intelligence (AI) or Machine Learning (ML) output in OGC-compliant geospatial data formats.

A GeoCollaborate workflow involves launching an instance and providing followers or dashboard users with a URL. A dashboard has many different interpretations by many different people. For GeoCollaborate, a dashboard means that anyone with the URL can access a map with the shared data layers that the leader has authorized. These data layers can be turned on or off per the leader's instructions. If data layers from the collaboration instance are turned off, they will not appear on the dashboard. If the data layer is turned off in the 'Session' layer then it will still appear in the dashboard but not be displayed. This enables a subject matter expert to 'activate' or turn on specific data layers that are of higher priority for a particular hazard or incident that is being tracked. The users of the dashboard can immediately see the data layers that are important and which ones may be helpful. Each user of the dashboard can turn on and off their own data layers based on their desires to see more information.

GeoCollaborate also has a 'Key Points' window that can be used as a one way information delivery window to keep collaborators informed of updates, new data layers being added, when a briefing is scheduled to occur, or specific information about ARD or DRI data layers that can be delivered to assist in data use and interpretation. GeoCollaborate is the technology designed to connect people together for collaborative decision making rapidly which is a goal of DP21. During the August Complex wildfire in 2020 in California, GeoCollaborate was used to coordinate the CA National Guard and the CA Civil Air Patrol (CAP) to identify imaging targets that were likely damaged or destroyed by the wildfire but needed to be confirmed by the CA CAP. The only way to image these facilities is to fly over them and take pictures. Heavy smoke was preventing these missions along with a temporary flight restriction (TFR) by the FAA over the wildfire location. GeoCollaborate was used to bring all critical assets together from disparate servers into one collaborative common operating picture (C-COP) along with the NOAA HRRR (High Resolution Rapid Refresh) smoke model to identify when it would be clear to fly. This cross-platform data sharing in a collaborative environment enabled all key decision makers to be on the same map at the same time while looking at the same data. The following brief video (Length 7:37), demonstrated how GC was used in this particular use case.

5.12.5. Technology interchange experiments

GeoCollaborate is designed to access and share data to accelerate situational awareness and decision making. While not demonstrated in this project, GeoCollaborate is designed to integrate into other mapping systems to enable collaborative data sharing, even across products. This capability enables cross-product data sharing and interoperability. There are other use cases for real-time geospatial data sharing, for example, GeoCollaborate is being funded under a technology innovation fund grant from the State of Florida to improve the access and sharing of data related to harmful algal blooms in the Indian River Lagoon (IRL) on the East coast of the state. The work being done in this OGC DP21 effort may lead to the addition of Decision Ready Indicators (DRI) into the GeoCollaborate instances serving the IRL. A demonstration of how GeoCollaborate is being used in this project can be accessed here: <u>Video Demonstration of GC in Indian River Lagoon</u> Length 7:20

5.12.6. Challenges and lessons learned

5.12.6.1. Challenge 1. Identifying consistent web services

While the pilot did develop data layers that were served via OGC-compliant web services, those services were not always working since they were being tested and updated fairly consistently. This provided occasional challenges when GeoCollaborate was being demonstrated with DP21 data layers. When the service is not working, the data layers will not be accessed or shared. Perhaps future pilots could provide a standard location to host services more consistently so demonstrations could be provided at any time and the up-time percentage of those servers could be close to 100%. It is to be expected that web services will be updated frequently during a Disasters Pilot. This is where innovation happens and is much more experimental than operational.

5.12.6.2. Challenge 2. More work needs to be done to assess the value of the GeoPackage data format so it is friendly for collaboration.

While GeoPackage provides a secure way to share data and package up certain sensitive data for sharing, there are also other 'secure' environments evolving that could provide a rapid 'cloaked' capability to initiate collaboration sessions that are invisible to any non-invited collaboration participant. These cloaked sessions could be very attractive to organizations in both the civil, defense and intelligence sectors and use cases could be developed that look at the value of creating a 'cloaked' environment vs. generating GeoPackages that keep data and/or pieces of sensitive data secure for sharing. When rapid collaboration sessions need to be initiated, establishing an entirely 'cloaked' environment may be mush easier than making sure all data providers are using the GeoPackage data format.

5.12.6.3. Challenge 3. Getting sponsors & stakeholders to participate in technology demonstrations

Stakeholder meetings are valuable so participants can report on progress made during the OGC Pilot. It is much more valuable when the stakeholders/sponsors actually participate actively by asking questions and inviting innovation that can directly impact their goals. When presentations are given and no feedback is delivered by the stakeholders and/or sponsors, the ability to move forward is limited by simply executing on the funded proposal. More progress could be shown with heightened interactivity.

5.12.6.4. Challenge 4. Engagement of end users during pilot projects

If future pilot project can work to engage users ahead of time and keep them engaged through the pilot, StormCenter believes that the implementation and success rates of OGC pilots will increase. It would be good to see what the end goal is of the sponsors so participants can deliver on those goals or at least make progress toward those goals. Engaging the end users can go a long way in making this successful for everyone. Sponsors would like solutions and participants would like sponsors to use/purchase or license their technologies. That is a win/win/win for everyone.

5.12.6.5. Challenge 5. Telling a comprehensive story about how all elements of DP21 (and future Pilots) serve the sponsors

There were so many excellent activities going on within the DP21 project it was tough to communicate how those pieces fit together to deliver a pathway forward for sponsors and users. Identifying gaps and areas of future activities will power future Disaster Pilots that will benefit sponsors, users, and members of OGC. It would also be good to understand how the DP21 and future pilots fit into the sponsor's plans and objectives. StormCenter is confident that this will evolve as future pilots get funded and executed.

5.12.7. Lessons learned

This was the very first project where StormCenter participated as a funded OGC participant. There was a lot to learn including how to navigate and publish reports and updates using the web portal and Wiki while also learning the basics of Asciidoc publishing. The weekly meetings were good to stay engaged and provide updates. It might have been beneficial to place some of the participants in direct communication with the sponsors or have sponsor-specific meetings to discuss progress and whether or not they had questions about the progress of the project or of any individual activity being funded. Some sponsors were hard to connect with and others seemed to keep 'arms-length' distance, which may prevent the open exchange of information that could guide progress.

The development of the 'Guides' is an excellent step forward to make data producers and data users better understand how to think about producing and using trusted data sets to drive decision making.

5.12.8. Next steps: updates and applications

DP21 produced an instance of GeoCollaborate that was used for demonstrations throughout the DP21 pilot. This instance is available until June 2022 which is six months after the end of DP21. The 'Live' DP21 Dashboard is available 24/7/365 and training is required to operate the Lead instance. The image below shows the DP21 dashboard with an HSR.Health Flood Layer and alternate route to use to get to medical facilities.

GeoCollaborate can play a significant role in connecting OGC pilot data services, including ARD and DRI, into decision making environments. This could introduce decision makers to the significance of web-based data services, OGC data Standards, and the burden it removes from decision makers when data is assigned specific Operational Readiness Levels (ORLs). ORLs have been developed by the ESIP Federation and the All-Hazards Consortium (AHC) to provide a level of confidence in data quality and trust to accelerate situational awareness and decision making by non-science experts. While much attention is given to the systems, algorithms, and processes to produce ARD and DRI, not enough attention is given to engaging the end users and decision-makers so the data can be put to use effectively and efficiently.

NOTEIf the data layers served by other DP21 participants do not appear when clicked (turning them on), the web service may not be operating at the moment. Some data layers load slowly due to the server being used in this test pilot.

Figure 88 — The GeoCollaborate DP21 dashboard reflects any data layers that have been activated by an authorized Leader. Also pictured are two small graphics beside data layers named 'N43 Plan 1' and 'N43 Plan 2'. These ORL1 datasets are provided by NOAA and identify flight plans for future NOAA missions into tropical cyclones. The logo next to the dataset is an Operational Readiness Level (ORL1) indicator.

StormCenter can train OGC personnel on using GeCollaborate or participate in additional demonstrations with OGC and potential sponsors. This could be a very good way to show how the pieces of DP21 come together into a collaborative common operating picture to improve situational awareness and decision-making.
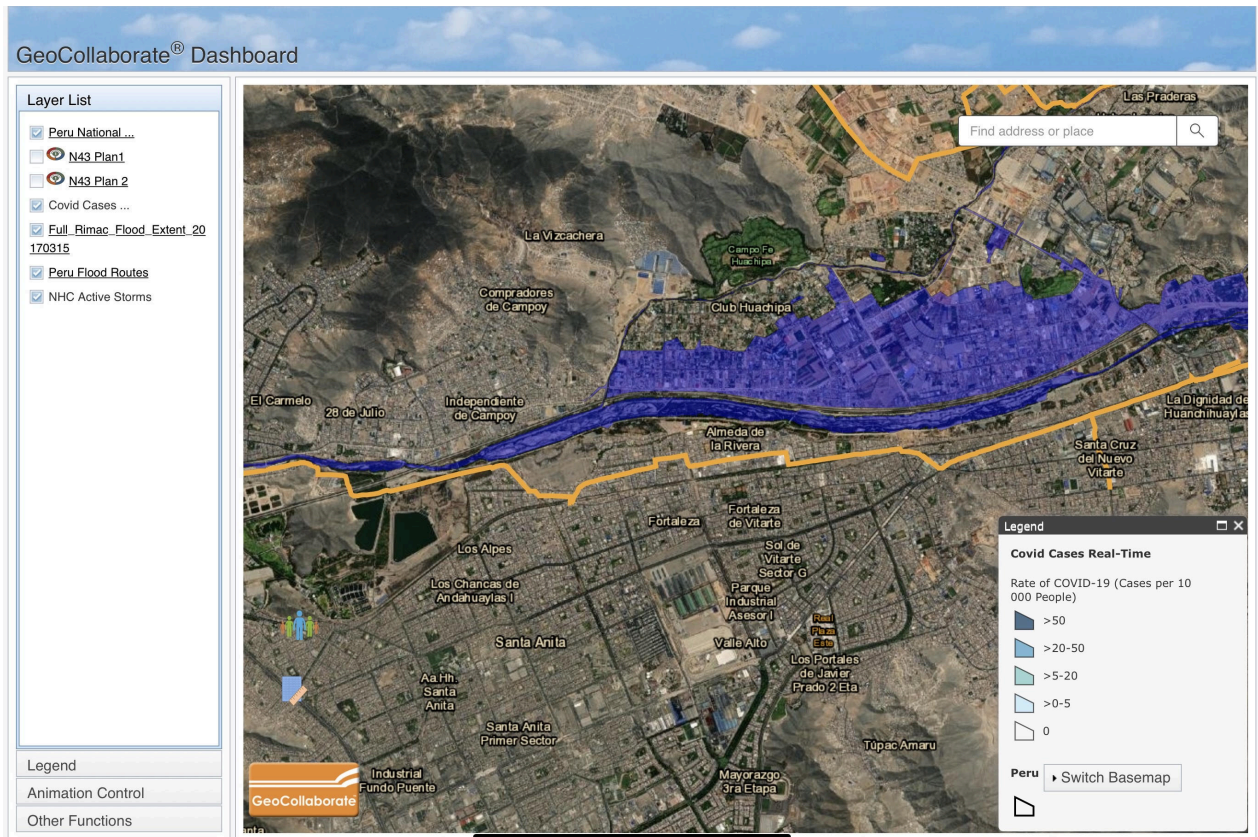
Figure 89 — The GeoCollaborate DP21 dashboard displayed on an iPad Pro and is connected to the collaboration session LIVE with data layers that have been activated by an authorized Leader.

5.13. EO Value Stream Component (Terradue)

5.13.1. EO Value Stream

For almost four decades, "Earth Observation" (EO) satellites developed and/or operated by Space Agencies have provided a wealth of data. In the past years, the Sentinel missions, along with the Copernicus Contributing Missions, Earth Explorers, and new commercial missions provide new ways to perform the routine monitoring of our environment at the global scale delivering an unprecedented amount of data. This expanding operational capability of global monitoring from space, combined with data from long-term EO archives (e.g., ERS, Envisat, Landsat), in-situ networks and models provide users with unprecedented insight into how

our oceans, atmosphere, land, and ice operate and interact as part of an interconnected Earth System.

However, while the availability of the growing volume of environmental data from space represents a unique opportunity for science, general R&D, and applications, it also poses a major challenge to achieve its full potential in terms of data exploitation focusing on the development of downstream products and services for the whole Disaster Reduction and Response cycle.

Firstly, because the emergence of large volumes of data (Petabytes era) raises new issues in terms of discovery, access, exploitation, and visualization of "Big Data," with profound implications on how users do "data-intensive" Earth Science. Secondly, because the inherent growing diversity and complexity of data and users, whereby different communities – having different needs, skills, methods, languages and protocols – need to cooperate to make sense of a wealth of data of different nature (e.g., EO, in-situ, model), structure, format, and error budgets.

Responding to these technological and community challenges requires the development of new ways of working, capitalizing on Information and Communication Technology (ICT) developments to facilitate the exploitation, analysis, sharing, mining, and visualization of massive EO data sets and high-level products within Europe and beyond. Evolution in information technology and the consequent shifts in user behavior and expectations provide new opportunities to provide more significant support to EO data exploitation.

However the Decision makers using Earth Observation data have very specific needs and the raw data coming from satellite missions cannot be exploited directly. EO products need to go through an EO value stream that derives the information contained in EO products and act as an interface between the technical characteristics of the satellite and the specific needs of the decision maker.

With the new capabilities brought by Earth Observation Exploitation Platforms, this EO-focused value stream flows from Acquisition to Decision, with new steps for product acquisition, ingestion, calibration and the actual production of Analysis Ready Data or even more importantly Decision Ready Information from multiple satellite missions taking full advantage of the new Cloud technologies.

As seen in next figure, at each step, different actors, interfaces, and components play an important role in five major facets focused on: Discovery, Content, Access, Format, and Visualization.
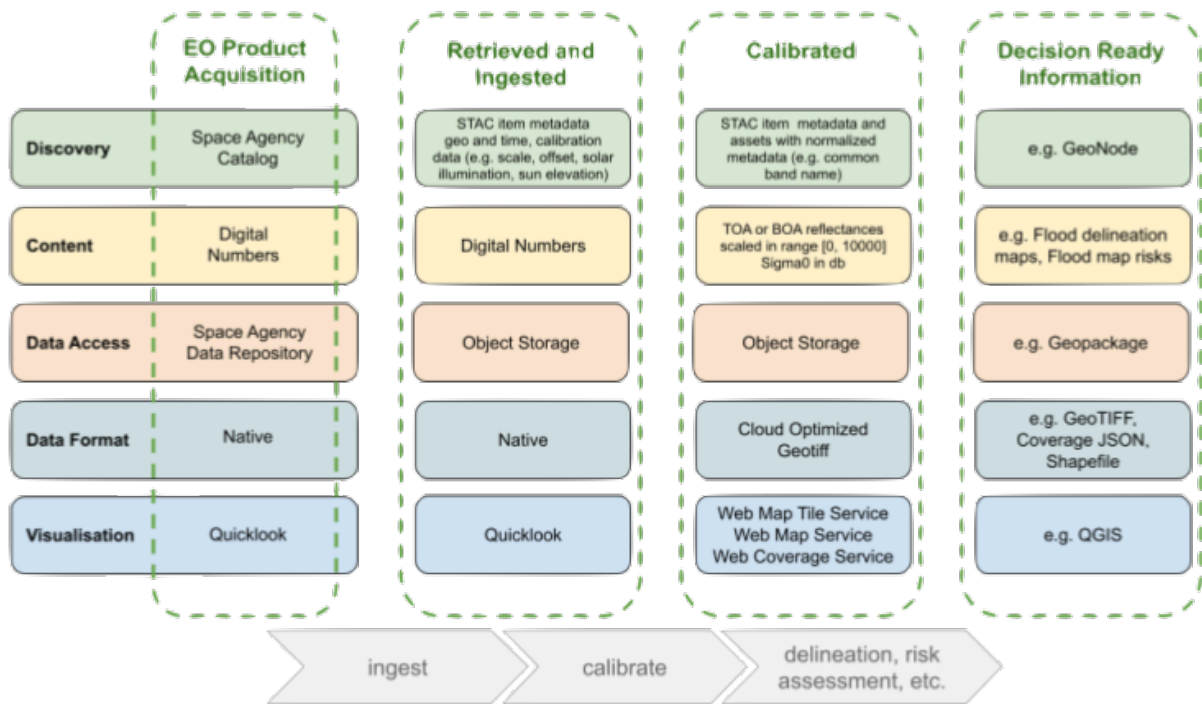
Figure 90 — EO value stream from Acquisition to Decision

At the EO product acquisition step, the data discovery and access are managed by the providers (e.g., Space Agencies) that provide their catalogs and repositories with the products in their native format and data content (e.g., digital numbers). The access methods and policies are heterogeneous from provider to provider and when available, the visualization is limited to quicklooks or previews.

In the next step the communities retrieved and ingested these products and even if the content and data format are still the same, the ingestion process can create specialized data discovery services. These discovery services specialization are focused on the enrichment of the metadata with information relevant for the community discovery process and their process workflow specialized for the domain in a catalog with metadata targeted for their needs. At this stage the data access is further facilitated by the usage of more Cloud friendly storage like Object Storage.

After the ingestion, an initial processing is performed to calibrate the EO data coming from multiple missions with the objective of making it ready for further generic processing. To achieve a multi-mission environment a calibration procedure is needed to enable a faster thematic processing of optical and SAR data. The optical calibrated products are pre-processed to obtain data with physical meaning with Top or Bottom of Atmosphere reflectances (TOA, BOA) and SAR calibrated datasets are pre-processed SAR detected images, given as sigma, beta, or gamma nought in dB.

At this calibration level the multi-mission products metadata are also normalized using Common Band Names (CBN) to ease the discovery process. CBN classes refer to common band ranges derived from predefined frequency ranges of the Electromagnetic Spectrum made for several popular instruments. Each CBN class is defined by a Band Range in micrometers for optical and in centimeters for SAR data. The CBN classification of the frequency spectrum allows a one-

to-one mapping of multi-mission and multi-sensor bands (Optical and SAR) and thus eases the handling of multi-sensor source EO data.

The calibration step is key because it significantly changes the data, format and visualization of the products. The data content are no longer digital numbers but actual BOA or TOA reflectances and are now stored as Cloud Optimized GeoTIFF (COG) and the visualization can now fully use the typical set of OGC Services (WTMS, WMS, WCS).

With the implementation of the previous steps in a process it is possible to perform the actual processing of multi-mission data to create Decision Ready Information. The adoption of the OGC Best Practice for Earth Observation Application Package (OGC 20-089) facilitates the direct deployment of specialized algorithms in a fast and responsive process responding to the needs of the actual scope of the disaster being addressed. At this stage the output data content is far from being an EO product but is now targeting the actual contextual information (e.g., flood delimitation maps) ready to be directly in the event. The information output is packaged (e.g., Geopackage) in common formats (e.g., GeoTiff, GeoJSON, Shape Files) ready to be used by desktop applications (e.g., QGIS).

The following sections describe the technical details of the EO value stream flow from Acquisition to Decision, with the new steps for product ingestion and calibration from multiple satellite missions taking full advantage of the new Cloud technologies. The section concludes with an overview of the selected Technical Components that were used to fulfill the deployment of those steps.

5.13.1.1. Ingestion

This ingestion process phase handles the retrieval and ingestion of EO acquisitions from the missions' ground segment to the platform storage and catalog.

The retrieval step foresees:

- The physical copy of the EO acquisition to an object storage exposing the S3 interface; and

- The generation of an EO acquisition manifest using the STAC Item specification including a minimum set of properties (e.g., mission, sensor, geo, and time) and a STAC asset with an href to the object storage location of the EO acquisition.

The next step in the ingestion process phase includes:

- Using metadata extractors against the retrieved EO acquisition; and

- Generating an extracted version of the EO acquisition with an extended STAC Item manifest including the STAC assets referring to all the EO acquisition data and metadata files.

At this stage, the STAC Item includes the EO acquisition properties to allow its optical or SAR calibration. For the optical calibration, the STAC Assets include elements allowing the transformation of the Digital Numbers to Top-of-Atmosphere or Bottom-of-Atmosphere

reflectances (scale, offset, solar illumination, and sun elevation). The SAR metadata elements include the incidence angle, the polarization among others.

The different facets of the Ingestion process phase are as follows.

- **Discovery**: The EO acquisition is first discovered using the Agencies' catalog services. Typically, the interfaces exposed are heterogeneous despite the standardization efforts and thus one of the goals of the discovery process phase is to expose the ingested EO acquisitions in a STAC catalog exposing the OGC API features interface. At this stage, the ingested EO acquisitions are searchable and discoverable thanks to the STAC Items properties.

- **Content**: During the ingestion phase, the EO acquisition content is not modified.

- **Data access**: The ingested EO acquisition is accessible via the object storage and its S3 interface.

- **Data format**: During the ingestion phase, the data format is not modified.

- **Visualization**: The visualization is limited to what the Agency included in the EO acquisition. It spans from nothing at all to simple JPEG or PNG images.

5.13.1.2. Calibration

The calibration process phase takes the ingested EO acquisition and applies a processing step generating "normalized" data which can be TOA or BOA reflectances for optical data and geocoded sigma0, beta0 or gamma0 for SAR data. The different facets of the Calibration processes are as follows.

- **Discovery**: The discovery is managed by a STAC catalog exposing the OGC API Features. Queries using CQL or CQL2 filters can be used to discover the calibrated EO acquisitions. The STAC Items' assets include metadata using the STAC extensions to support their discovery.

- **Content**: The content after the ingestion is not modified and typically represents raw data. After the calibration process phase, the content is now geocoded TOA or BOA reflectances and Sigma0, Beta0, or Gamma0 for SAR.

- **Data access**: The calibrated products are now optimized for remote access or conventional download as they're stored on object storage exposing the S3 interface.

- **Data format**: The data format is COG, a Cloud native format allowing optimized range access to portions of the data being radiometric or geographic.

- **Visualization**: The visualization of calibrated assets includes direct access to the calibrated products STAC Item and assets. This includes OGC interfaces like WMS, WMTS, or WMC.

5.13.2. Discovery Component

5.13.2.1. STAC API/OGC API Features

The Discovery component we chose for the Disaster Pilot 21 is the STAC FastAPI implementation of the STAC API/OGC API Features. The STAC FastAPI was deployed on a Kubernetes cluster and exposed on the Internet.

Collections can be created to group entries according to common entry characteristics or business goals. In this context, we chose to create collections for the retrieved, ingested and calibrated EO acquisitions.

Below the "calibrated" collection JSON payload:

```
{
  "type": "Collection",
  "id": "calibrated",
  "stac_version": "1.0.0",
  "description": "OGC DP 21 Calibrated",
  "links": [
    {
      "rel": "root",
      "href": null,
      "type": "application/json"
    }
  ],
  "stac_extensions": [],
  "extent": {
    "spatial": {
      "bbox": [ -180, -90, 180, 90]
    },
    "temporal": {
      "interval": [
        ["2020-01-01T00:00:00Z",null]
      ]
    }
  },
  "license": "proprietary"
}
```

The REST endpoint exposed allows inserting catalog entries by posting STAC items and running queries by posting CQL2 filters encoded in JSON.

Below is an example of a STAC Item of a calibrated EO product from RADARSAT Constellation Mission (RCM) that was provided during the activities of the Pilot:

```
{
  "type": "Feature",
  "stac_version": "1.0.0",
  "id": "RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_HH_HV_GRD",
  "properties": {
    "datetime": "2020-04-16T12:47:37Z",
    "created": "2020-04-16T12:47:37Z",
    "updated": "2022-01-25T01:20:19Z",
    "mission": "rcm",
    "platform": "rcm-3",
```

```
        "instruments": [ "sar" ],
        "title": "Rcm-3 GRD HH/HV",
        "sat:orbit_state": "descending",
        "sar:instrument_mode": "",
        "sar:frequency_band": "C",
        "sar:polarizations": [ "HH", "HV" ],
        "sar:product_type": "GRD",
        "sar:observation_direction": "right",
        "sar:center_frequency": 5.404999999449389,
        "sar:resolution_range": 0,
        "sar:resolution_azimuth": 0,
        "sar:pixel_spacing_range": 2,
        "sar:pixel_spacing_azimuth": 2,
        "sar:looks_range": 1,
        "sar:looks_azimuth": 1,
        "view:incidence_angle": 29.379944801330566
    },
    "geometry": {
        "type": "Polygon",
        "coordinates": [
            [
                [-97.319422038811, 48.795372425465],
                [-96.908770730352, 48.741913175672
              ],
                [ -96.761332124828, 49.217979194055],
                [ -97.176114807754, 49.271647243182],
                [ -97.319422038811, 48.795372425465]
            ]
        ]
    },
    "links": [
        {
            "rel": "self", "type": "application/geo+json",
            "href": "s3://ogc-dp21/rcm-calibrated/RCM3_OK1070382_PK1074766_1_
5M6_20200416_124737_HH_HV_GRD/RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_
HH_HV_GRD.json"
        },
        {
            "rel": "root", "type": "application/json",
            "href": "s3://ogc-dp21/rcm-calibrated/catalog.json"
        }
    ],
    "assets": {
        "overview-HH": {
            "href": "s3://ogc-dp21/rcm-calibrated/RCM3_OK1070382_PK1074766_1_
5M6_20200416_124737_HH_HV_GRD/RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_
HH_HV_GRD_HH_overview.tif",
            "type": "image/tiff; application=geotiff",
            "title": "overview for sigma0 in decibel for polarization HH",
            "file:size": 105323709,
            "roles": [ "overview" ]
        },
        "s0-db-HH": {
            "href": "s3://ogc-dp21/rcm-calibrated/RCM3_OK1070382_PK1074766_1_
5M6_20200416_124737_HH_HV_GRD/RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_
HH_HV_GRD_HH.tif",
            "type": "image/tiff; application=geotiff",
            "title": "sigma0 in decibel for polarization HH",
            "file:size": 779101425,
            "roles": [ "data" ]
        },
        "overview-HV": {
```

```
            "href": "s3://ogc-dp21/rcm-calibrated/RCM3_OK1070382_PK1074766_1_
5M6_20200416_124737_HH_HV_GRD/RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_
HH_HV_GRD_HV_overview.tif",
            "type": "image/tiff; application=geotiff",
            "title": "overview for sigma0 in decibel for polarization HV",
            "file:size": 92450539,
            "roles": [ "overview" ]
        },
        "s0-db-HV": {
            "href": "s3://ogc-dp21/rcm-calibrated/RCM3_OK1070382_PK1074766_1_
5M6_20200416_124737_HH_HV_GRD/RCM3_OK1070382_PK1074766_1_5M6_20200416_124737_
HH_HV_GRD_HV.tif",
            "type": "image/tiff; application=geotiff",
            "title": "sigma0 in decibel for polarization HV",
            "file:size": 779101425,
            "roles": [ "data" ]
        }
    },
    "bbox": [-97.319422038811, 48.741913175672, -96.761332124828, 49.
271647243182 ],
    "stac_extensions": [
        "https://stac-extensions.github.io/processing/v1.0.0/schema.json",
        "https://stac-extensions.github.io/projection/v1.0.0/schema.json",
        "https://stac-extensions.github.io/sar/v1.0.0/schema.json",
        "https://stac-extensions.github.io/sat/v1.0.0/schema.json",
        "https://stac-extensions.github.io/view/v1.0.0/schema.json"
    ],
    "collection": "calibrated"
}
```

CQL2 filters can be used to query the catalog. Below an example of such filter in Python:

```
daterange = {"interval": ["2020-04-01T00:00:00Z", "2020-04-29T23:59:59Z"]}

filt = {
  "op": "and",
  "args": [
    {"op": "=", "args": [{"property": "collection"}, "calibrated"]},
    {"op": "=", "args": [{"property": "sat:orbit_state"}, "descending"]},
    {"op": "anyinteracts", "args": [{"property": "datetime"}, daterange]},
  ]
}
```

5.13.3. Access Component

5.13.3.1. Object Storage (S3)

The access to calibrated EO acquisitions is backed by object storage exposing the S3 interface. S3 is part of the offer of several Cloud providers or can be installed and managed with open source implementations like MinIO.

The tooling in the S3 Object Storage ecosystem is quite rich and the exploitation of objects stored on S3 using Python is reasonably simple and quite powerful. Listing or accessing the contents of stored objects is reasonably straightforward.

Below a few code snippets used in the demonstration notebooks allow the exploration of the Object Storage contents.

```python
S3Settings = namedtuple(
    "S3Settings",
    ["region_name", "endpoint_url", "aws_access_key_id", "aws_secret_access_key"],
    defaults=[os.environ["S3_REGION"], os.environ["S3_SERVICE_URL"].replace("https://", ""), None, None],
)

secret_key = os.environ["S3_SECRET_KEY"]
access_key = os.environ["S3_ACCESS_KEY"]

settings = S3Settings(
    aws_access_key_id=access_key, aws_secret_access_key=secret_key
)

session = botocore.session.Session()

s3 = session.create_client(
    service_name="s3",
    region_name=settings.region_name,
    use_ssl=True,
    endpoint_url=f"https://{settings.endpoint_url}",
    aws_access_key_id=settings.aws_access_key_id,
    aws_secret_access_key=settings.aws_secret_access_key,
)

bucket = "ogc-dp21"
folder = "rcm-calibrated"
paginator = s3.get_paginator("list_objects_v2")
pages = paginator.paginate(Bucket=bucket)

item_urls = []

for page in pages:
    item_urls.extend(
        [
            f"s3://{bucket}/" + obj["Key"]
            for obj in page["Contents"]
            if obj["Key"].endswith(".json") and 'catalog' not in obj["Key"]
and folder in obj["Key"]
        ]
    )
```

This returns (clipped):

```
["s3://ogc-dp21/rcm-calibrated/RCM1_OK1721243_.../..._PK1721268_4_5M9_20200425_
001248_HH_HV_GRD.json",
 "s3://ogc-dp21/rcm-calibrated/RCM1_OK1727158_.../..._PK1727289_4_5M15_
20200414_002048_HH_HV_GRD.json",
 ...
 "s3://ogc-dp21/rcm-calibrated/RCM3_OK1719525_.../..._PK1719577_4_5M15_
20200422_002111_HH_HV_GRD.json",
 "s3://ogc-dp21/rcm-calibrated/RCM3_OK1743118_.../..._PK1743174_3_5M9_20200421_
001311_HH_HV_GRD.json",
 "s3://ogc-dp21/rcm-calibrated/RCM3_OK1743118_.../..._PK1743175_3_5M15_
20200410_002110_HH_HV_GRD.json"]
```

Below, accessing a COG file with GDAL:

```python
gdal.SetConfigOption('AWS_REGION', os.environ["S3_REGION"])
```

```
gdal.SetConfigOption('AWS_SECRET_ACCESS_KEY', os.environ["S3_SECRET_KEY"])
gdal.SetConfigOption('AWS_ACCESS_KEY_ID', os.environ["S3_ACCESS_KEY"])
gdal.SetConfigOption('AWS_S3_ENDPOINT', 's3.fr-par.scw.cloud')
gdal.SetConfigOption('AWS_S3_ENDPOINT', os.environ["S3_SERVICE_URL"].
replace("https://", ""))

gdal_url = item.get_assets()["s0-db-HV"].get_absolute_href().replace("s3://",
"/vsis3/")

ds = gdal.Open(gdal_url)
band = ds.GetRasterBand(1).ReadAsArray(1000,1000,1000,1000)
band
```

The band variable contains:

```
array([[  0., 0., 0.  , ..., -14.46626   , -10.833421, -11.105058 ],
       [  0., 0., 0.  , ..., -12.656098  , -9.090598, -15.903594 ],
       [  0., 0.,   0.  , ..., -12.936415  , -6.215864, -17.451353 ], ...,
       [  0.49788436, -6.1458416, -8.018673, ..., -13.702349, -17.989513, -14.
56787 ],
       [ -4.189909, -4.761789, -11.1462, ..., -21.63922, -14.363132, -20.
343006 ],
       [ -7.967144, -2.1650217, -10.353745, ..., -18.561003,-13.033768, -12.
409282 ]],
       dtype=float32)
```

5.13.4. Processing Component

The proposed definition for an EO Application package according to OGC 20-089r1 is:

A platform independent and self-contained representation of an Application, providing executables, metadata and dependencies such that it can be deployed to and executed within an Exploitation Platform.

The Application Package is a document that describes the data processing application by providing information about the parameters, software item, executable, dependencies, and metadata. This file document ensures that the application is fully portable among all supporting processing scenarios and supports automatic deployment in a Machine-To-Machine (M2M) scenario. Most importantly, the Application Package information model allows the deployment of the application as an OGC API — Processes (OGC 18-062) compliant web service.

The Application Package includes the following information:

- Reference to the executable block that implements the Application functionality; and

- Description of its input/output interface.

The Application Package uses the Common Workflow Language (CWL) Workflow Description specification as an encoding to describe the Application, its parameters, command-line tools, their runtime environments, their arguments, and their invocation within containers.

During the Disaster Pilot, several application packages and associated containers were developed. Below we list a few that handled the RCM EO acquisitions made available during the Pilot.

- **Retrieval** : This EO application package takes a local RCM acquisition and implements the acquisition minimal metadata extraction and upload to S3. The outcome is then a registered RCM acquisition in the "retrieved" collection.

- **Ingestion** : This EO application package takes a local RCM acquisition and implements the acquisition metadata extraction by dedicated metadata extractors and upload to S3. The outcome is then a registered RCM acquisition in the "ingested" collection.

- **Calibration** : This EO application package takes an ingested RCM acquisition and implements the SAR calibration and upload to S3. The outcome is then a registered RCM calibrated acquisition in the "calibrated" collection.

The application packages are then exposed on a platform that:

- Accepts a Post request with an Application Package (OGC API - Processes);

- Translates Application Package metadata to create a new process in the OGC API - Processes instance;

- Translates Application Package Workflow Inputs defined in the CWL document as OGC API - Processes parameters; and

- Creates a new Process offering in the OGC API - Processes.

At this stage, the platform can accept execution requests and, for each request, it:

- Translates OGC API - Processes execute parameters to the Workflow Inputs defined in the Application Package (CWL document);

- If applicable, executes the data stage-in for the input EO products;

- Orchestrates and executes CWL; and

- Translates output to OGC API — Processes outputs.

We used the Earth Observation Exploitation Platforms Common Architecture (EOEPCA) implementation of the OGC API Processes that combines the Zoo project for the interface and a Kubernetes based processing backend relying on the Calrissian project.

5.13.5. Visualization Component

The visualization component selected is Titiler, a stateless service accepting as input a reference to a COG file or to a STAC Item and returning a OGC WMTS endpoint (among other options). In our perspective, this component encapsulates a pure Cloud native visualization service for the Pilot's calibrated data as it can read the STAC Items as inputs that have STAC assets pointing to S3 URLs.

Below a code snippet that requests an overview of one RCM STAC Item asset:

```
titiler_endpoint = "https://tiler.ogc-disaster-pilot-21.terradue.com"
```

```
service = "cog/preview"

url = urllib.parse.quote(pairs[1]['pre_event'].get_assets()["s0-db-HV"].get_
absolute_href())

response = requests.get(f'{titiler_endpoint}/{service}?format=png&url=
{url}&max_size=512&resampling_method=nearest&rescale=-26%2C0&return_mask=true')
img = Image.open(BytesIO(response.content))
```

In a Jupyter notebook, this is the output:

Figure 91 — EO value stream Visualization Component using Titiler in a Jupyter Notebook

5.13.6. Future Work

During the activities performed in this pilot we were able to define the EO value stream flow to support decision makers to use Earth Observation. The EO value stream flow derives the information contained in EO products and acts as an interface between the technical characteristics of the satellite and the specific needs of the decision maker. Taking advantage of the Earth Observation Exploitation Platforms and the available Cloud technologies, the flow goes from Acquisition to Decision. It defines steps for product acquisition, ingestion, calibration, and the actual production of Analysis Ready Data or even more importantly Decision Ready Information from multiple satellite missions.

In the activities performed we identify the Technical Components necessary for the product ingestion and calibration steps of the EO value stream flow. Further work is necessary

to identify and demonstrate the full flow reaching the production of the Decision Ready Information. In particular, further activity is necessary to extend the catalog (e.g., STAC) with the necessary information model ready to support the whole Disaster Reduction and Response cycle. Conversely one key point that was identified is that even if reproducibility is most surely provided by the Application Package it is necessary to provide exact lineage that links the information produced with the original EO products.

From the Disaster Response perspective we also identified the need for a process to identify and generate reference datasets during a disaster. For example, while responding to a flood event we are able to identify the necessary dataset and processing to obtain the exact flood extent but not so straightforward to provide the community with the reference situation previously or not under the influence of the event that they are addressing.

5.14. Analysis Ready Data Component (Wuhan University)

5.14.1. About Wuhan University

Wuhan University (WHU) is a university that plays a significant role in researching and teaching all aspects of surveying and mapping, remote sensing, photogrammetry, and geospatial information sciences in China. The School of Remote Sensing and Information Engineering, WHU, is a research center dedicated to developing standard-based geospatial information technologies and applications in the Earth science domain. Our group has been working on geospatial interoperability, data processing, provenance, and Machine Learning for many years and has wide influence among the Chinese GIS community. The group members have rich experiences in developing standards and technologies for geospatial interoperability, such as the participation in OGC innovation programs including OWS-9, OWS-10, Testbed 12, Testbed13, Testbed 17, Disaster Resilience Pilot 2019 (DRP-2019), Environmental Data Retrieval (EDR) API SWG, and Training Data Markup Language for Artificial Intelligence (TrainingDML-AI) SWG. Following on the successes of DRP-2019 (User Guide, Demo Video), in Disaster Pilot 2021, we contributed to ARD Component, DRI Component, and Web/Desktop Analysis Component. We collaborated with other participants to show how independent components using standards can accomplish the pilot tasks and the scenario requirements, thus turning the collaborative success into a production pipeline for the disaster SDI ecosystem.

5.14.2. Description of the contributed components and role in the Pilot architecture

WHU participated in three components, as shown in Figure 92.
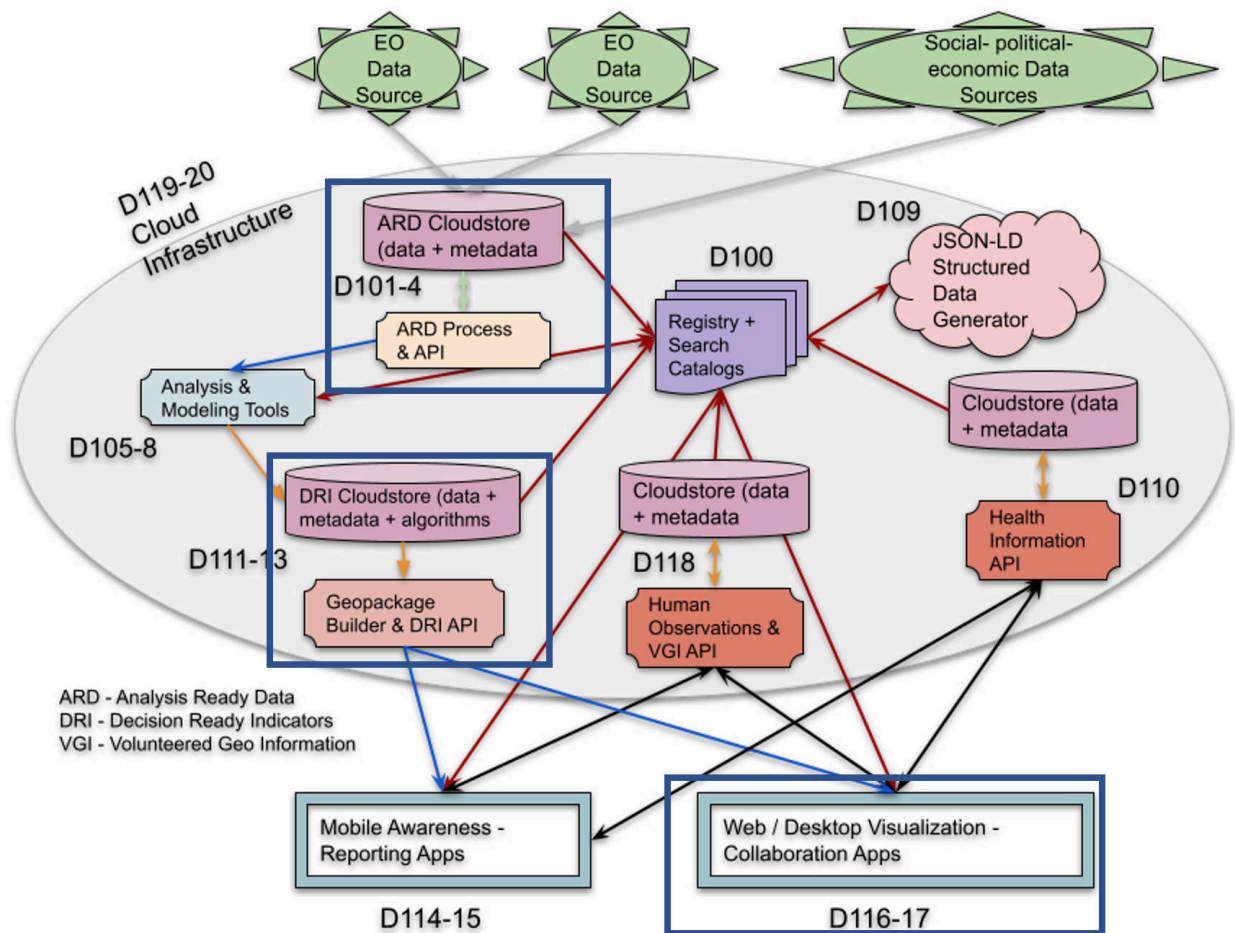
Figure 92 — Components that WHU participated in

- D101-4 Analysis Ready Data (ARD) Component: WHU collected multiple spatial-temporal datasets from open providers and DP21 stakeholders. These data were processed into the form that allows immediate flood analysis, then published as Web Services for the sharing with other components.

- D111-3 Decision Ready Information (DRI) Components: WHU built a DRI Cloud store based on MinIO to upload and download DRI data. WHU developed GeoPackage conversion tools for converting and packaging DRI data, including the SHP file, CSV file, and raster file in Int 8, Int16, or float32 formats.

- D116-7 Analysis, Visualization, Collaboration Components: WHU developed a resource center to register, manage, and query the diverse range of DRI and ARD data; WHU developed a browser-based visualization client that can discover, request, download, and visualization both ARD and DRI products for analysis and decision.

5.14.3. Introduction

By some estimates, 80% of response time is spent procuring and preparing data, only 20% can be spent learning from it. Therefore, Analysis Ready Data (ARD) is vital for responders to work

with the disaster analysis and support decision-making immediately. However, as OGC pointed out, *data, particularly EO data, can be hard to find, complicated to share, difficult to access, and slow (or unable) to be processed into common forms that are suitable for analysis and integration*. The goal of WHU is to develop a framework that allows rapid access to ARD in disaster response.

The most recognized definition of ARD is from CEOS, i.e., *satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.* In practice, they are usually Surface Reflection for optical imagery and Normalised Radar Backscatter for SAR imagery. However, the *minimum of additional user effort* varies from different analysis tasks. In DP21, WHU consider all the data with observation characteristics as ARD, which need. For example, Surface Reflection is ARD for flood detection, while flood inundation is ARD for further disaster assessment.

5.14.4. Technical design and implementation

Figure 93 shows the technical route on how WHU contributed to the ARD component. WHU collected open EO data from open EO providers (USGS, ESA, and NASA) and non-open EO data from DP21 stakeholders (NRCan, CONIDA, Satellogic). The advantage of the former is the free and open policy, while the advantage of the latter is the higher spatial resolution for precise monitoring. WHU developed a framework to discover, search and in parallel, download multi-source datasets by wrapping the Web API of open EO provider and GIS provider. After preprocessing, WHU gets the Top of Atmosphere (TOA) reflection or surface reflection of optical data and backscatter intensity of SAR data. These data are processed into flood products. WHU also collects essential geographical data like DEM, road network, landcover, and surface water to support impact assessment. EO-based ARD and geographical ARD support DRI analysis collaboratively.
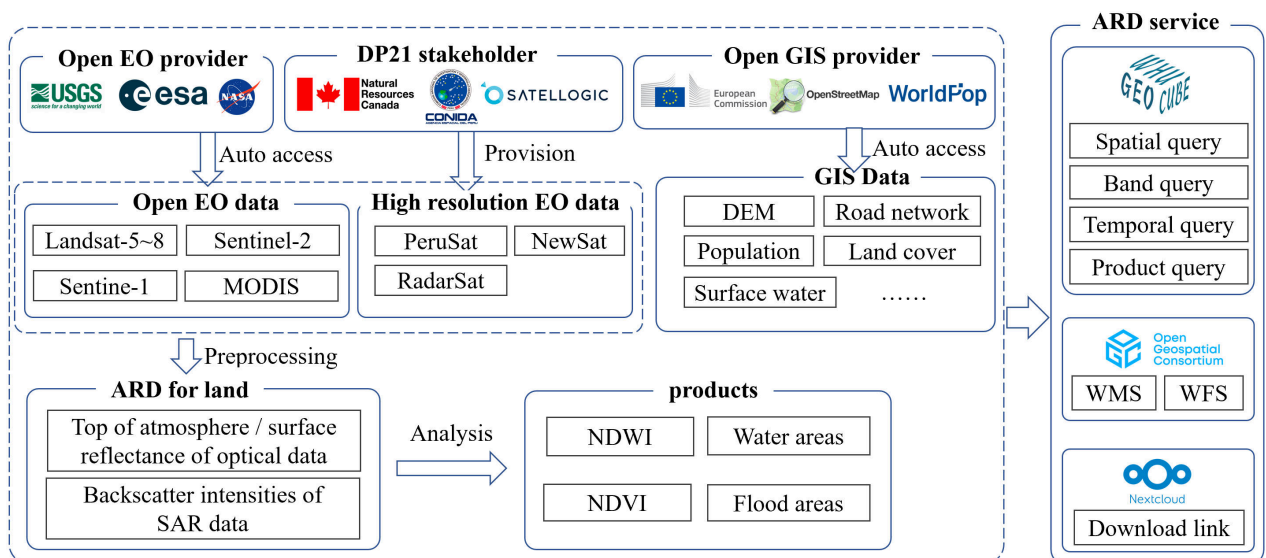
Figure 93 — Technical architecture of ARD component

Recipe for detecting water areas from optical imagery:

- Preprocessing

- NDWI

- Binarization

- Differencing

- Filtering

Recipe for detecting water areas from SAR imagery:

- Preprocessing

- Thresholding or supervised classification

- Binarization

- Differencing

- Filtering

WHU provided three types of ARD Services to other components.

- Geospatial data cube: WHU integrated NewSat imagery to GeoCube that supports n-dimensional semantic query, including spatial query, temporal query, band query, product query, etc.

- OGC Web Service: WHU published Raster ARD and Vector ARD as WMS and WFS, respectively, allowing seamless access in other standard-based components.

- Download link: WHU also uploaded these files to its Nextcloud storage to provide direct URLs. All WHU produced ARD for the Red River case are listed at https://docs.google.com/spreadsheets/d/1NnaQ82P1wCtnsc5hk3HC42zIMJZvcJeBcBtzz56b0-g. (Restricted by the policy of WHU's network, it is not accessible at midnight, Chinese Standard Time).

5.14.5. Case study 1: Flood event in the Red River basin

For the flood event in the Red River Basin in April 2020, we select a rectangular study area (95.94°W98.28°W, 47.48°N~ 49.84°N), as shown in Figure 94.

Figure 94 — Study area of the Red River case

Table 8 — EO ARD of the Red River case

DATA	PROVIDER	NUMBER OF SCENES	RESOLUTION
RCM	NRCan	10	5m
Sentinel-1	ASF DAAC	16	10m
Sentinel-2	ESA	18	10m
Landsat-8	USGS	12	30m
NewSat	Satellogic	9	1m

Table 9 — Precipitation ARD of the Red River case

DATA	PROVIDER	NUMBER OF SCENES	TEMPORAL RESOLUTION	SPATIAL RESOLUTION
GPM	NASA GES DISC	30	daily	0.1 degree

Table 10 — Geographical ARD of the Red River case

DATA	SOURCE	RESOLUTION
DEM	SRTM	30m
Land cover	FROM-GLC	30m
Population	WorldPop	90m
Road network	OSM	/
Slope	/	30m
HAND	/	30m
Water occurrence	JRC	30m

Table 11 — Flood ARD of the Red River case

DATE	SOURCE DATA	SOURCE DATA PROVIDER	RECIPE
10th, April	RCM sigma0	NRCan	Supervised classification
11th, April	Sentinel-1 GRD sigma0	ASF DAAC	thresholding
14th, April	RCM sigma0	NRCan	Supervised classification
16th, April	RCM sigma0	NRCan	Supervised classification
16th, April	Sentinel-1 GRD sigma0	ASF DAAC	thresholding
17th, April	Landsat-8 collection2 level-2	USGS	NDWI
17th, April	RCM sigma0	NRCan	Supervised classification
17th, April	Sentinel-2 level 2A	ESA	NDWI
21th, April	RCM sigma0	NRCan	Supervised classification
22th, April	Sentinel-2 level 2A	ESA	NDWI
23th, April	RCM sigma0	NRCan	Supervised classification

DATE	SOURCE DATA	SOURCE DATA PROVIDER	RECIPE
24th, April	Landsat-8 collection2 level-2	USGS	NDWI
25th, April	RCM sigma0	NRCan	Supervised classification
28th, April	RCM sigma0	NRCan	Supervised classification
28th, April	Sentinel-1 GRD sigma0	ASF DAAC	NDWI
30th, April	RCM sigma0	NRCan	Supervised classification

The following four figures shows some typical geographical ARD products.
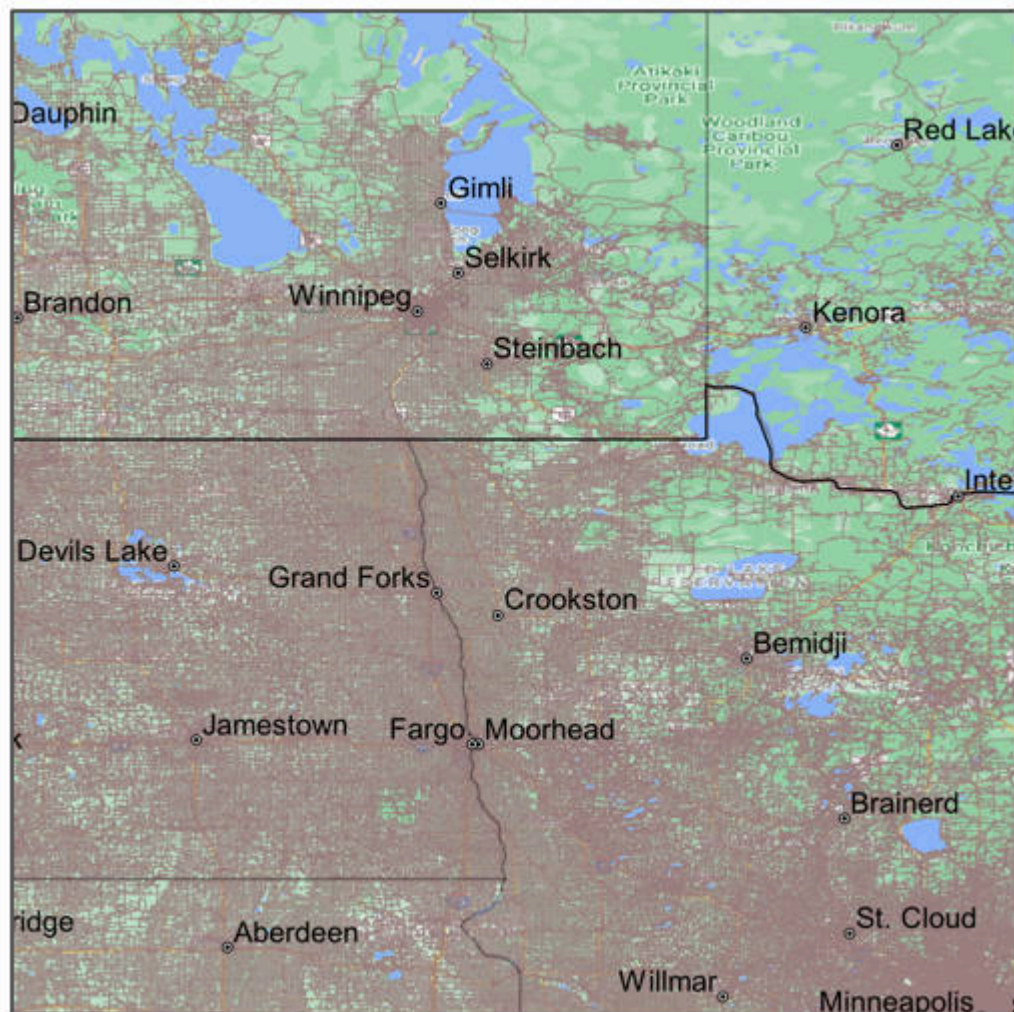
Figure 95 — Road network of the Red River case
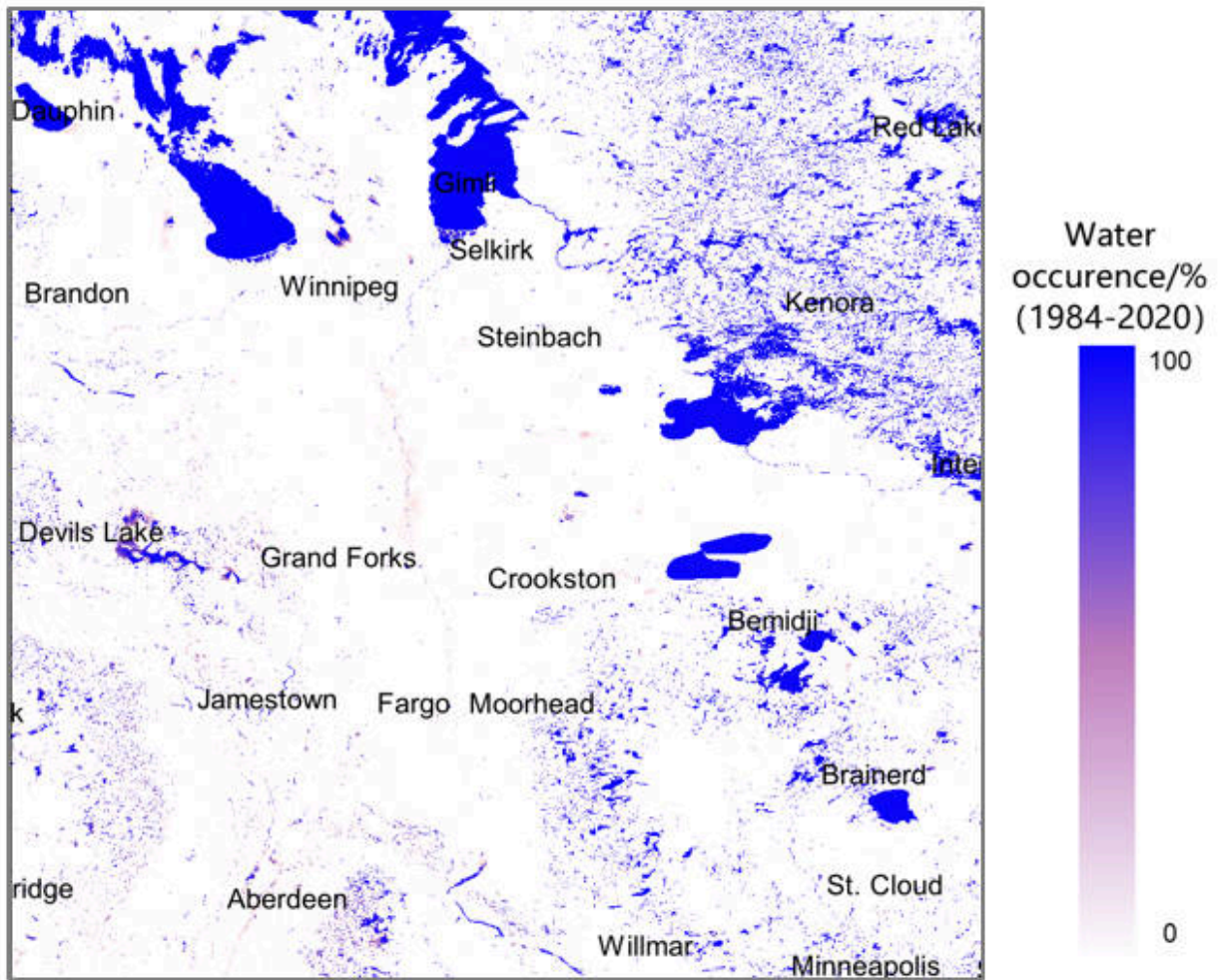
Figure 96 — Water occurrence of the Red River case
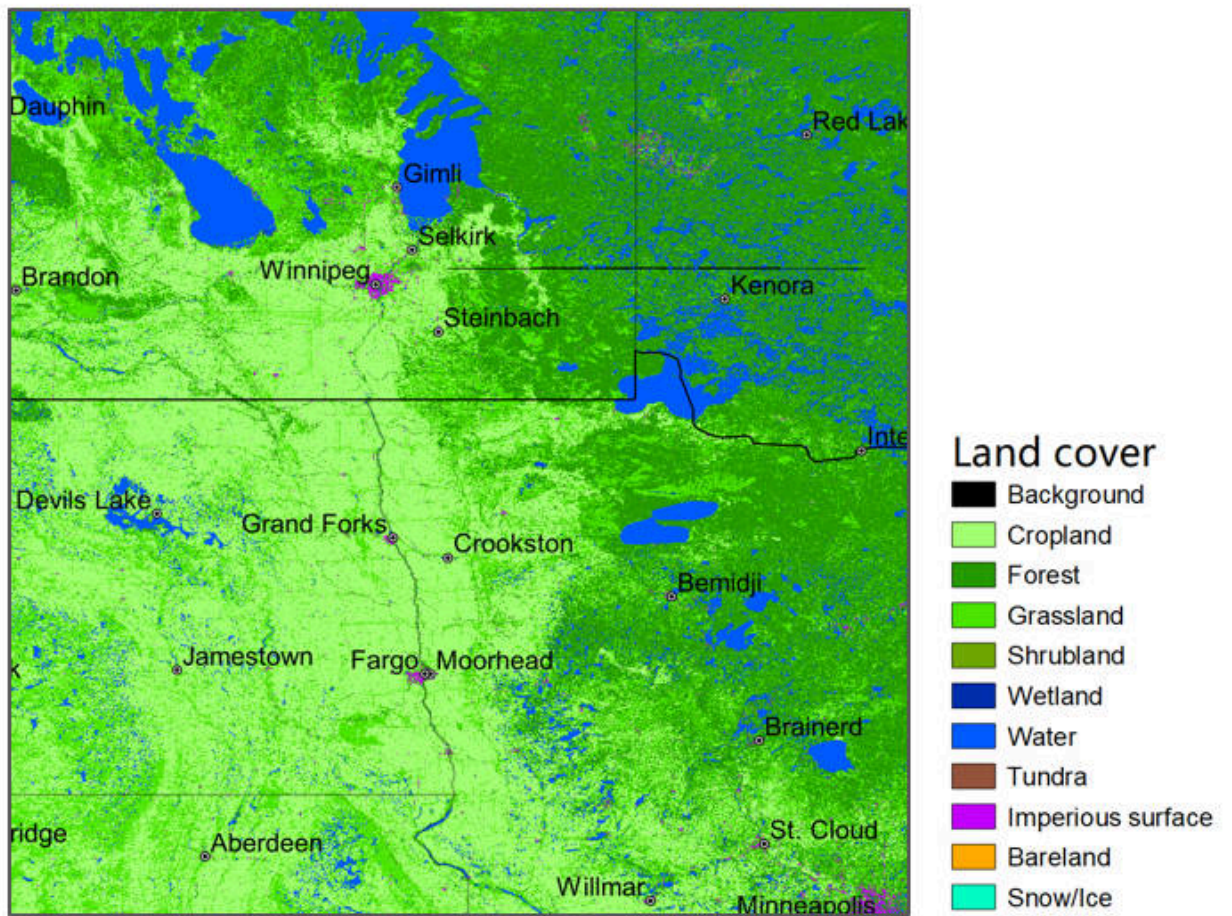
Figure 97 — Land cover of the Red River case
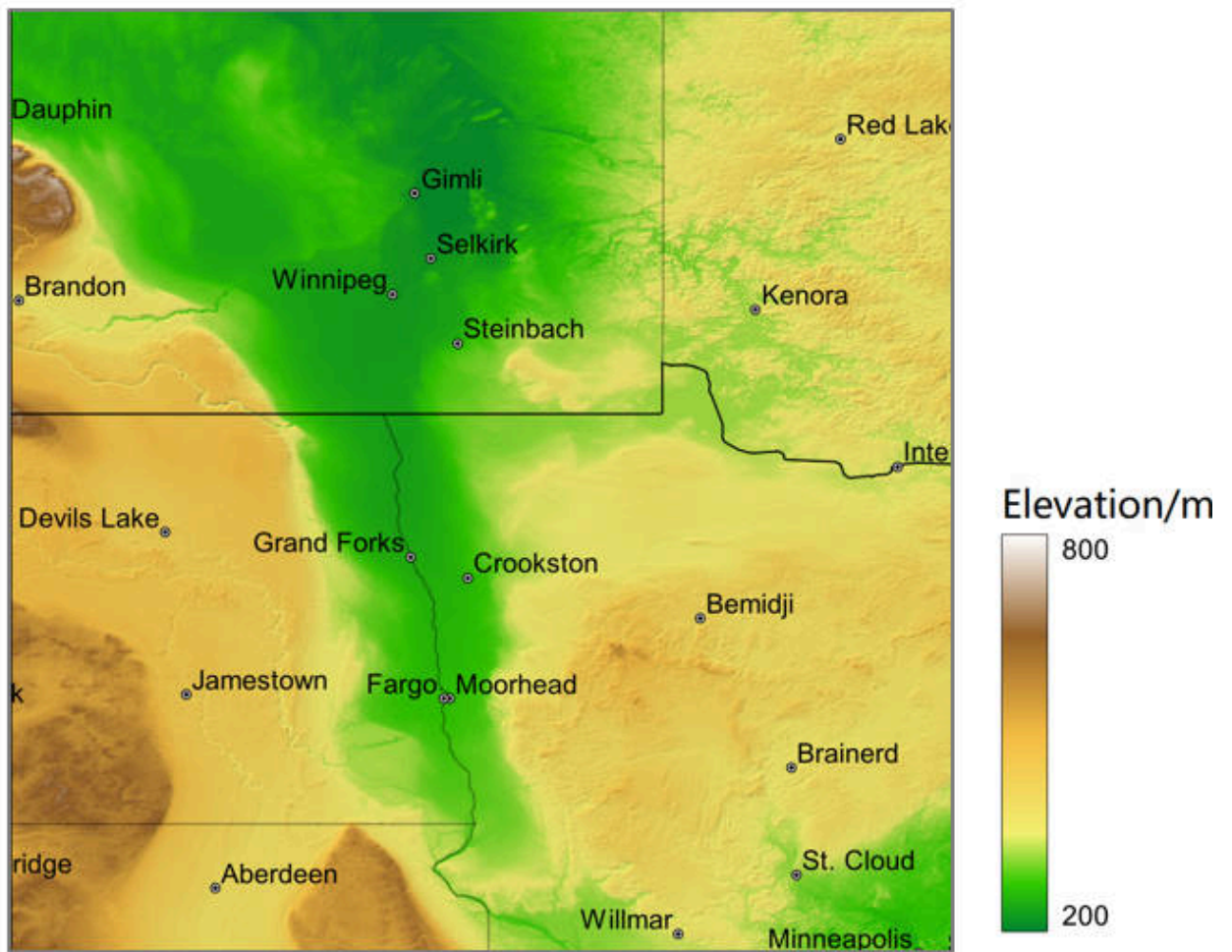
Figure 98 — DEM of the Red River case

The next three figures show the Time-series ARD products of water areas of the Red River case derived from multi-source datasets (RCM, Landsat-8, and Sentinel-1).
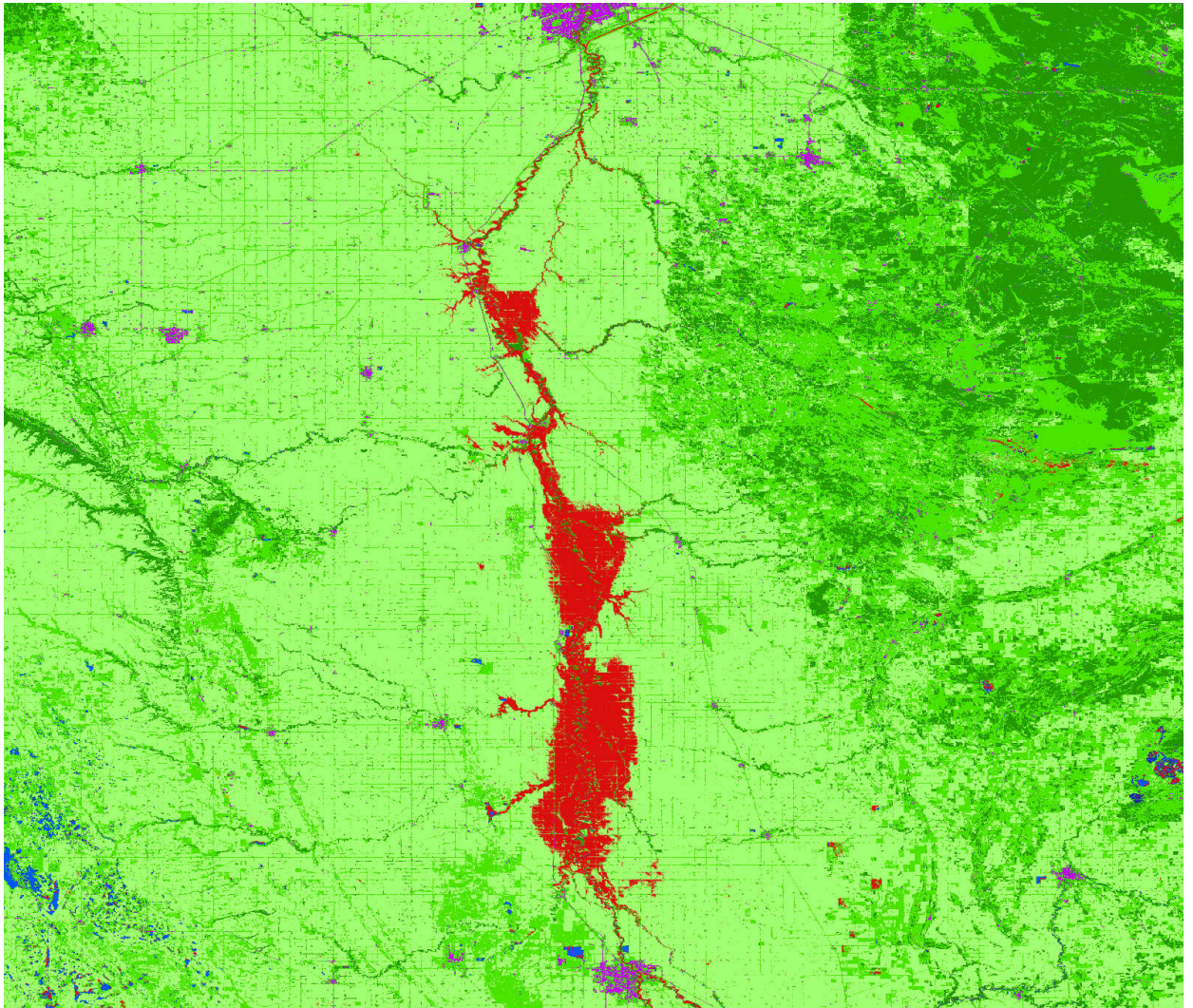
Figure 99 — Water areas of the Red River case on 17th, April, derived from Landsat-8
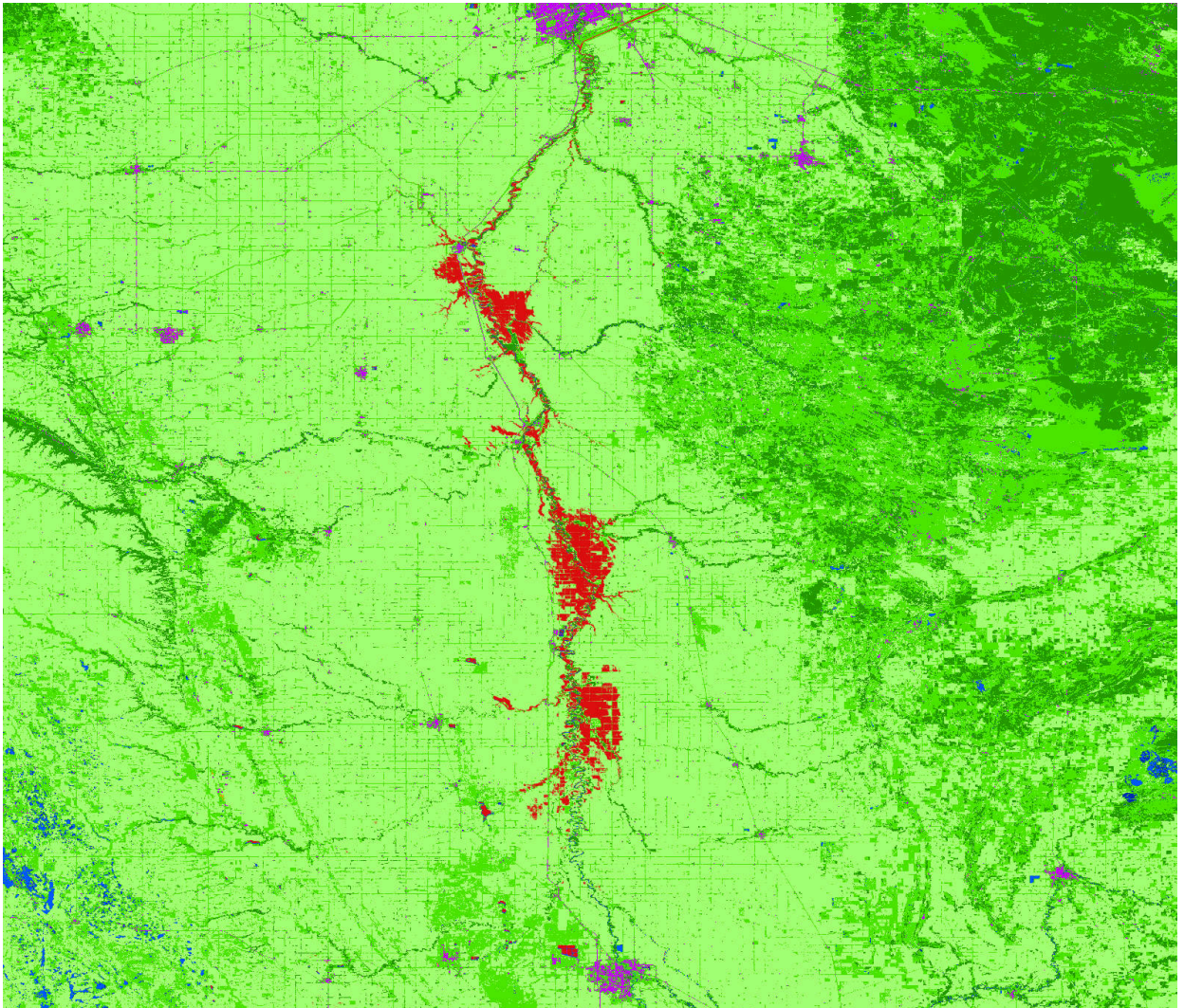
Figure 100 — Water areas of the Red River case on 30th April, derived from Sentinel-1
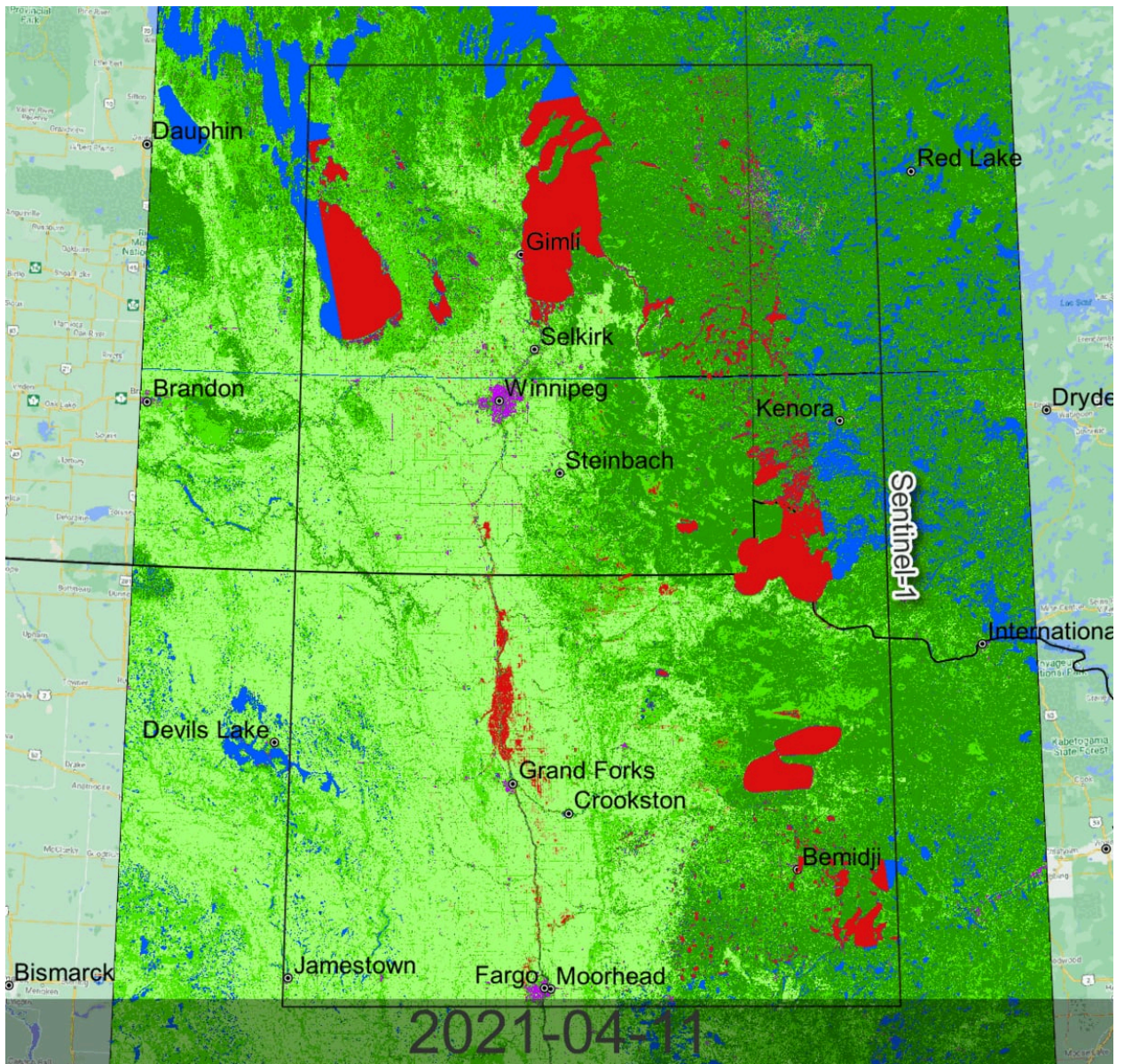
Figure 101 — Water areas of the Red River case on 30th April, derived from Sentinel-1

5.14.6. Case study 2: flood event in the Rímac and Piura river basin

For the flood event in the Rímac and Piura river basin in March 2021, we select Piura Province as study area, as shown in Figure 102.
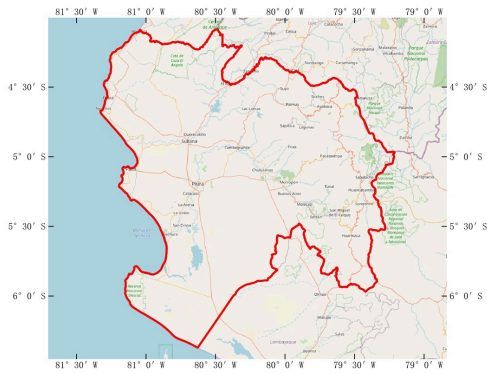
Figure 102 — Study area of Piura case

The next four tables list the ARD products WHU produced.

Table 12 — EO ARD of the Piura case

DATA	PROVIDER	NUMBER OF SCENES	RESOLUTION
RCM	NRCan	10	5m
Sentinel-1	ASF DAAC	16	10m
Sentinel-2	ESA	18	10m
Landsat-8	USGS	12	30m
NewSat	Satellogic	9	1m

Table 13 — Precipitation ARD of the Piura case

DATA	PROVIDER	NUMBER OF SCENES	TEMPORAL RESOLUTION	SPATIAL RESOLUTION
GPM	NASA GES DISC	31	daily	0.1 degree

Table 14 — Geographical ARD of the Piura case

DATA	SOURCE	RESOLUTION
DEM	SRTM	30m
Land cover	FROM-GLC	30m

DATA	SOURCE	RESOLUTION
Population	WorldPop	90m
Road network	OSM	/
Slope	/	30m
HAND	/	30m
Water occurrence	JRC	30m

Table 15 — Flood ARD of the Piura case

DATE	SOURCE DATA	SOURCE DATA PROVIDER	RECIPE
5th, March	Sentinel-2 level 2A	ESA	NDWI
7th, March	Sentinel-2 level 2A	ESA	NDWI
12th, March	Sentinel-2 level 2A	ESA	NDWI
14th, March	Landsat-8 collection2 level-2	USGS	NDWI
17th, March	Sentinel-2 level 2A	ESA	NDWI
20th, March	Sentinel-2 level 2A	ESA	NDWI
22th, March	Landsat-7 collection2 level-2	USGS	NDWI
22th, March	Sentinel-2 level 2A	ESA	NDWI
23th, March	Landsat-8 collection2 level-2	USGS	NDWI
25th, March	Sentinel-2 level 2A	ESA	NDWI
27th, March	Sentinel-2 level 2A	ESA	NDWI
30th, March	Landsat-8 collection2 level-2	USGS	NDWI
30th, March	Sentinel-2 level 2A	ESA	NDWI

The following three figures show some typical geographical ARD products.
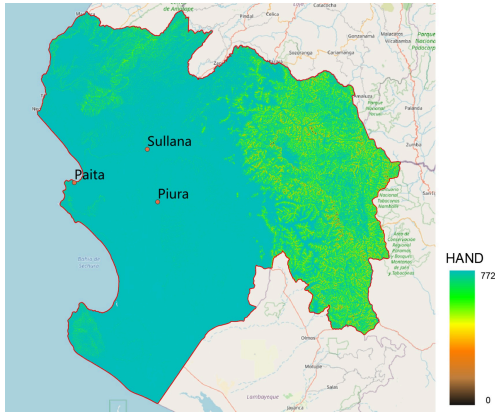
Figure 103 — HAND of Piura case
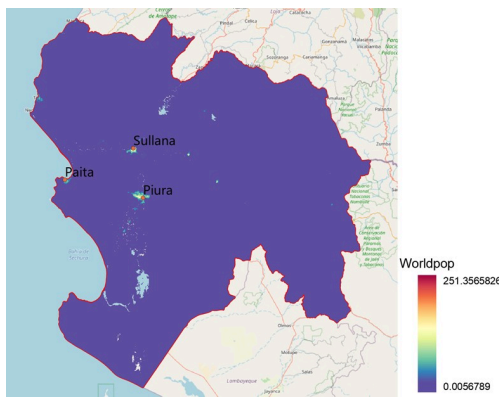
Figure 104 — Population of Piura case
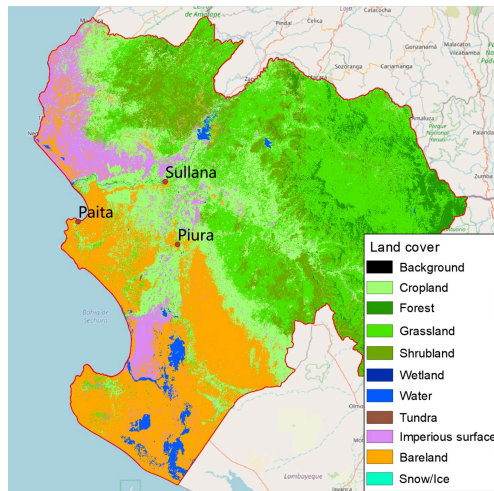
Figure 105 — Landcover of Piura case

The next four figures show the Time-series ARD products of water areas derived from multi-source datasets (Landsat-8, Landsat-7, and Sentinel-2).
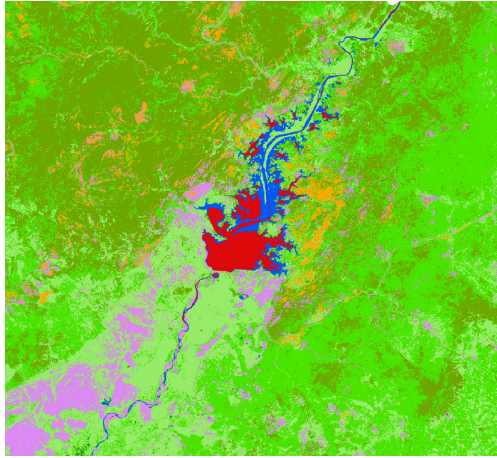
Figure 106 — Water areas of the Piura case on 10th, February, derived from Landsat-8
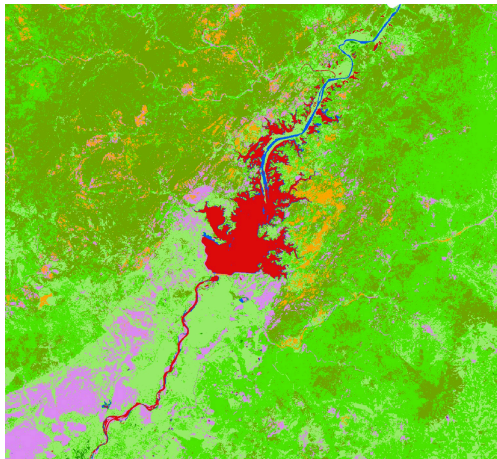
Figure 107 — Water areas of the Piura case on 14th, March, derived from Landsat-8
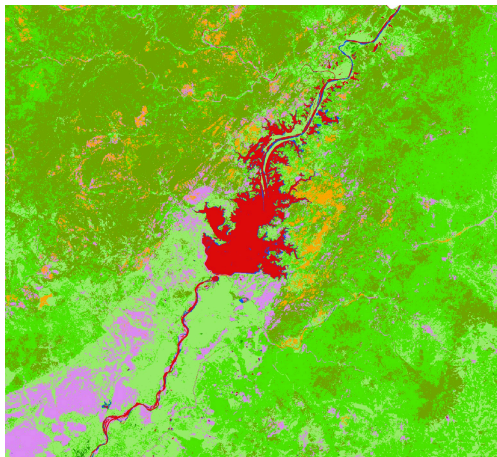
Figure 108 — Water areas of the Piura case on 22th, March, derived from Sentinel-2
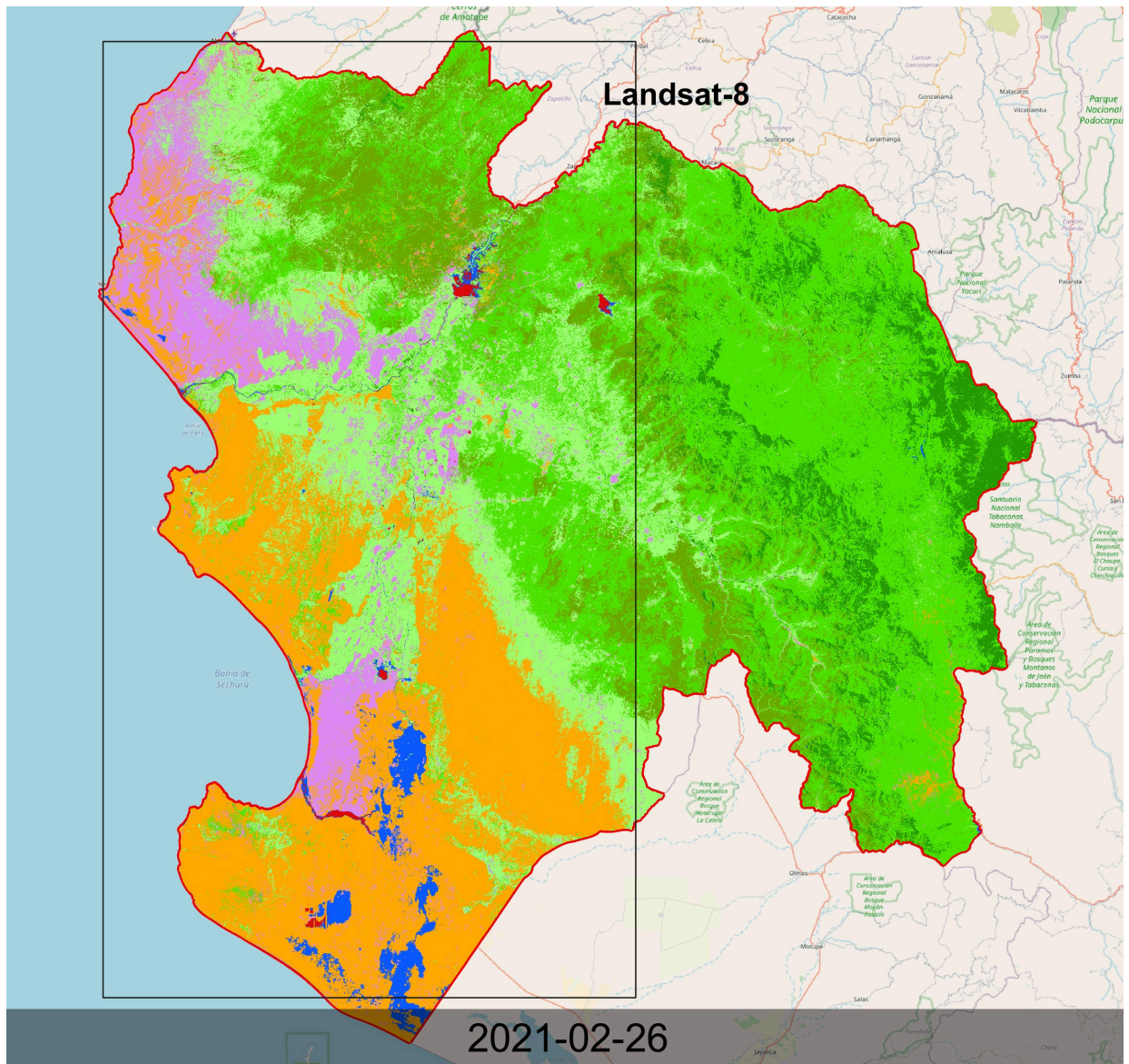
Figure 109 — Time-series ARD products of water areas of the Piura case from 26th February to 27th March, derived from multi-source datasets (Landsat-8, Landsat-7, and Sentinel-1)

5.15. Decision Ready Information Components (Wuhan University)

5.15.1. Introduction

After specific analysis, massive multi-source heterogeneous data can provide decision-makers with rich and valuable knowledge. However, this information is often so complex that it is

difficult for decision-makers and suffering people to find helpful information on time. As mentioned in Pilot's challenges, local information such as in situ sensor observations, field reports, and volunteered information are often difficult to collect and even more difficult to incorporate back into provided information products. Up-to-date and actionable event information, even when openly available, can be difficult for the affected public to find and stay on top of. Therefore, Decision Ready Information (DRI) Components was born.

Decision readiness information is a targeted data product for decision-makers and field responders that can be directly used in decision making. It is generally generated under the guidance of specific decision readiness indicators. It is necessary to consider the characteristics of multi-source heterogeneity of geospatial data, complexity and diversity of geo-analysis models, quality and reliability of decision readiness information products to construct decision readiness indicators rapidly and generate DRI products efficiently for ensuring the high efficiency of decision making and implementation. DRI Components can provide targeted information products to local analysts and field responders through modern convenience APIs, optimized hybrid-cloud services, and mobile-ready online-offline GeoPackage tools.

5.15.2. Technical design and implementation

(1) Technical solutions
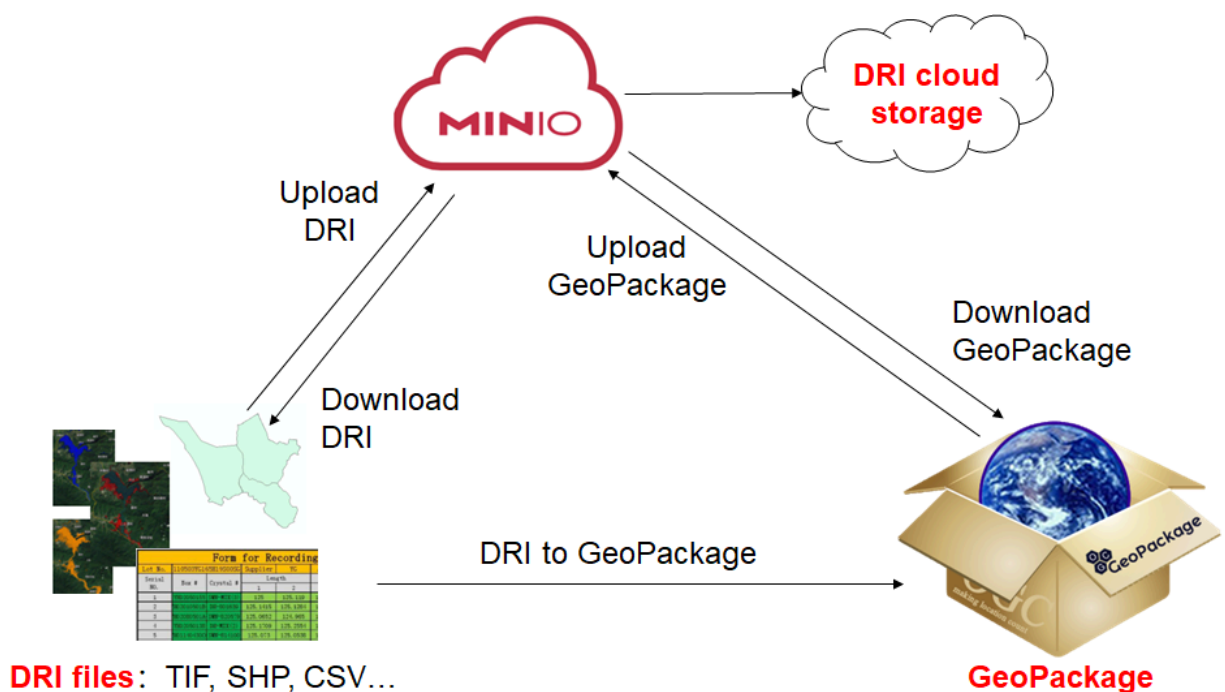
Figure 110 — Technical architecture of DRI Component

Figure 19 shows the technical architecture of the DRI Component. WHU provided a high-performance object storage service based on MinIO, connecting the upstream DRI files and the downstream GeoPackage files. First of all, When obtaining the upstream DRI files, WHU created two branches for storing DRI files and converting and packaging DRI files to GeoPackage files. Next, WHU provided unified delivery services by uploading GeoPackage files to the MinIO in a unified format. In the end, all users can upload and download the DRI files that they need

based on the convenient API provided by WHU. Under this mechanism, service and packaging elements are deployed "near" to decision-ready information products to make them discoverable and accessible to field application clients in environments of widely variable connectivity. This component can support the delivery of requested and/or subscribed information products as packages to address connected-disconnected operations.

(2) DRI cloud storage service

WHU provided DRI product data and metadata storage services as a data infrastructure to support the rapid construction of decision-ready indicators based on the MinIO High-Performance Object Storage Service.

The MinIO is an open-source, high-performance distributed object cloud storage service that implements most of the cloud storage service interfaces in the Amazon S3 standard. Compared with traditional HDFS, MinIO is simpler and more efficient, and it is good at processing files of any size (0-5TB). It adopts the separation mechanism of storage and computing, which is faster than Hadoop.

At present, WHU completed the DRI cloud storage service, including DRI file upload, DRI file download, GeoPackage file upload, and GeoPackage file download. It supports multiple files upload simultaneously.

The RESTful API of DRI data uploading is as follows： Post: http://125.220.153.26:8081/uploadDRI Parameter: "file", "parentPath".

The DRI data downloading API is as follows： Get: http://125.220.153.26:8081/downloadDRI Parameter: "filename".

(3) DRI product unified delivery services

WHU provided unified delivery services by converting all DRI data to GeoPackage data for effectively supporting field operations in disaster management situations. Visualization of the right information at the right time using GeoPackage data containers, with the generation of GeoPackage offline containers to allow taking all relevant information into the field even given connectivity issues.

GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial. A GeoPackage is the SQLite container, and the GeoPackage Encoding Standard governs the rules and requirements of content stored in a GeoPackage container.

WHU developed GeoPackage conversion tools for converting and packaging DRI data based on the OGC GeoPackage standard. These tools support the converting and packaging of multiple format files, including the SHP file, CSV file, and raster file in Int 8, Int16, or float32 formats. WHU provided these GeoPackage files to field sponsors and decision-makers based on the high-performance DRI cloud storage service.

5.16. Analysis, Visualization, Collaboration Components (Wuhan University)

5.16.1. Technical design and implementation

The analysis, visualization, collaboration component is implemented in a disaster emergency service platform, which consists of a database backend, services publishing layer, and Web browser-based visualization client. The database backend utilizes MySQL and PostgreSQL databases to store large-scale disaster data with multiple data types such as digital elevation models (DEMs), optical satellite imageries, SAR imageries, thematic maps, traffic networks, and precipitation statistics.

To manage the diverse range of DRI and ARD data, WHU developed a resource center based on the Spring-boot framework. As shown in Figure 11-13, the center can register and query these data published by the standard-base service such as WMS, WMTS, and WFS. In order to obtain inundated areas from raw remote sensing imagery data, data access services, as well as processing services, are necessary. Data access to the backend database is realized by WCS and WFS services.

Finally, WHU developed a browser-based visualization client using Javascript and HTML. The client can discover, request, download, register, and visualization both ARD and DRI products for analysis and decision in different scenarios, such as result data display as well as comparison with other data. For instance, comparing flood inundated areas with population data and road networks enables further analysis of the flood damages on roads, population, etc. Web browsers to be supported and tested include Internet Explorer 11 and up, Microsoft Edge, Google Chrome 46 and up, Firefox 42 and up, and Safari 9 and up.

Figure 111 — The data search page of resources center

Figure 112 — The metadata page of resources center

Figure 113 — The data registration page of resources center
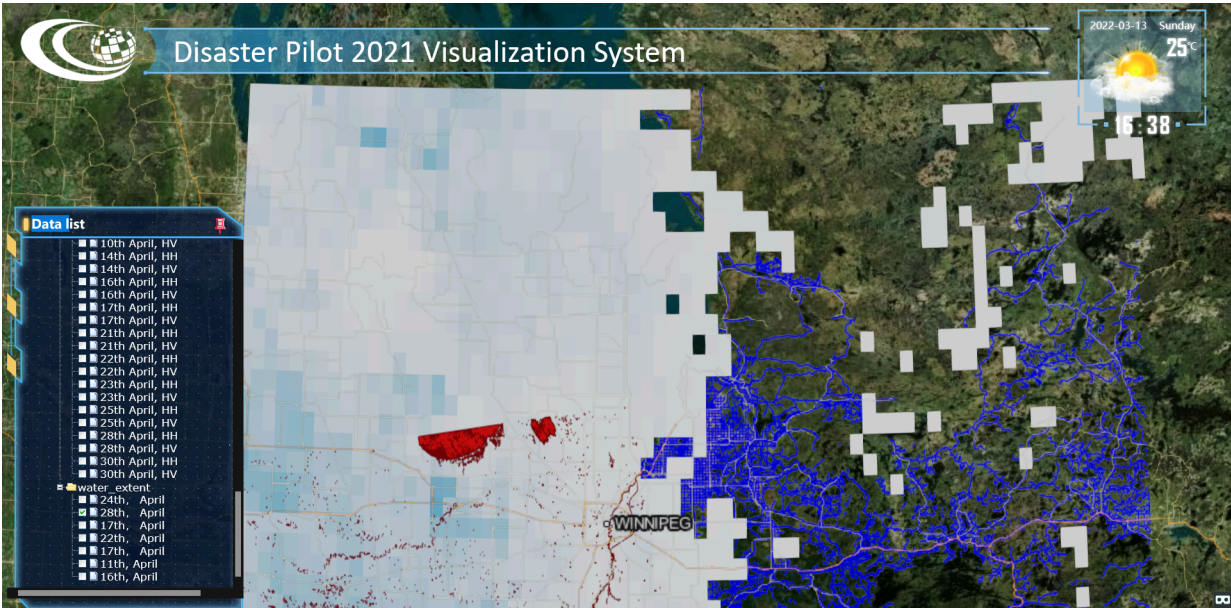
Figure 114 — Data Visualization: Users can view every piece of data by checking the corresponding box in the data list.

6

SCENARIOS

SCENARIOS

6.1. Peru Scenario

For the Peru Scenario HSR.health calculated its Medical Supply Needs Index (MSNI) for the Rimac River Basin. The MSNI calculates a range for the amount of medical supplies needed based upon the current prevalence and potential spread of the COVID-19 pandemic, the prevalence of at-risk populations, the current hospitalization rates, the number of medical personnel and first responders, and the burn rates of the supplies. One of the components of the Medical Supply Needs Index is the Pandemic Risk Index (PRI), which is comprised of the Transmission Risk Index (TRI) and the Mortality Risk Index (MRI).

The Mortality Risk Index utilizes data on population demographics and the prevalence of co-morbidities to identify the risk to the underlying population of severe illness, hospitalizations, or mortality due to the COVID-19 pandemic. The TRI utilizes data on population, case counts, area, and human mobility to identify the risk of spread of COVID-19 by geographic area. The PRI is a combination of these two indices which represents both the spread of the infectious disease and the health risk that the pandemic poses. These indices are extensible to other infectious diseases.

The MSNI produces three levels of need: high, low, and index-based levels of need. The high and low levels of medical supply needs are based on hospitalizations, ICU hospitalizations, supply utilization by type, in this case personal protective equipment (PPE) such as gowns, gloves, and masks, the number of healthcare workers and first responders such as police, firefighters, and EMS personnel. Once the high and low estimates are calculated the Pandemic Risk Index is incorporated to produce the supply level based upon the spread of the pandemic and the health risk to the underlying population.

The recommendation for action to the decision makers will be identified through calculating the difference between the Medical Supply Needs Index and the current supply stock at each hospital and clinic to determine when and how much supply needs to be delivered. With a coincident natural disaster, the medical supply needs calculation can be adjusted based upon increased volume of patients (due to the disaster) and based upon a shift in load on the medical systems if a hospital or clinic becomes inaccessible due to the natural disaster.

As a part of this pilot, HSR.health worked with Skymantics who are building a routing API that was utilized to identify the routes that should be taken to move the supplies from the distribution centers (such as warehouses, airports, or ports) to the hospitals and clinics. Skymantics' routing API provides dynamic routing information that allows for the identification of which roads are passable and impassable in a disaster scenario. Skymantics demonstrated this with a flooding scenario in the Rimac Basin and showed how the routes to the clinics changed due to the flooding scenario and how the supply needs changed as well.

The MSNI is housed on the GeoHealth Platform, which is a Health Spatial Data Infrastructure (SDI) built on a GeoNode and GeoServer infrastructure developed in collaboration with

Geosolutions. This is a geospatial application that can also be implemented on the ArcGIS Online infrastructure.

6.2. US Scenario

For the US Scenario HSR.health calculated its Medical Supply Needs Index (MSNI) for New Orleans, Louisiana at the Census tract level. The MSNI calculates a range for the amount of medical supplies needed based upon the current prevalence and potential spread of the COVID-19 pandemic, the prevalence of at-risk populations, the current hospitalization rates, the number of medical personnel and first responders, and the burn rates of the supplies. One of the components of the Medical Supply Needs Index is the Pandemic Risk Index (PRI), which is comprised of the Transmission Risk Index (TRI) and the Mortality Risk Index (MRI).

The Mortality Risk Index utilizes data on population demographics and the prevalence of co-morbidities to identify the risk to the underlying population of severe illness, hospitalizations, or mortality due to the COVID-19 pandemic. The TRI utilizes data on population, case counts, area, and human mobility to identify the risk of spread of COVID-19 by geographic area. The PRI is a combination of these two indices which represents both the spread of the infectious disease and the health risk that the pandemic poses. These indices are extensible to other infectious diseases.

The MSNI produces three levels of need: high, low, and index-based levels of need. The high and low levels of medical supply needs are based on hospitalizations, ICU hospitalizations, supply utilization by type, in this case personal protective equipment (PPE) such as gowns, gloves, and masks, the number of healthcare workers and first responders such as police, firefighters, and EMS personnel. Once the high and low estimates are calculated the Pandemic Risk Index is incorporated to produce the supply level based upon the spread of the pandemic and the health risk to the underlying population.

The recommendation for action to the decision makers will be identified through calculating the difference between the Medical Supply Needs Index and the current supply stock at each hospital and clinic to determine when and how much supply needs to be delivered. With a coincident natural disaster, the medical supply needs calculation can be adjusted based upon increased volume of patients (due to the disaster) and based upon a shift in load on the medical systems if a hospital or clinic becomes inaccessible due to the natural disaster.

The MSNI is housed on the GeoHealth Platform, which is a Health Spatial Data Infrastructure (SDI) built on a GeoNode and GeoServer infrastructure developed in collaboration with Geosolutions. This is a geospatial application that can also be implemented on the ArcGIS Online infrastructure.

Additionally, as a part of the US Scenario HSR.health built upon the existing Health SDI Concept Development Study by the OGC. This included conducting research into the availability of health data in the US for Louisiana and expanding upon the current data model. The three main areas of data and indicators that we expanded upon include:

- Essential Foundation Data and Technology Data — which includes base layers like road networks and hospital locations;

- Health Indicators for Disasters — such as identifying electricity dependent people for dialysis or ventilators and infected water supplies; and

- Health Indicators for Pandemics — such as case counts, deaths, mortality risk, vaccination rate, etc.

In order to ascertain local health issues, the Census tract level of geography is the ideal granularity. Louisiana has some pandemic related information such as case counts and deaths at the Census tract level which is one of the reasons why it was chosen for the US scenario. Several states do not publicly share pandemic data down to the Census tract or ZIP Code level. Data at the county or state levels is a starting point, but it limits the ability to do local level analyses of health issues. For the MSNI the social determinants data (sourced from the US Census Bureau) and case data (Louisiana State Department of Health) was available at the Census tract level. Prevalence rates for co-morbid health conditions are often available at the County and State level, and are less readily available at the Census tract level.

6.3. Canada Scenario

Through the sponsorship of Natural Resources Canada, the Pilot was able to look at hazards of flooding and pandemics in the portion of the Red River basin situated in the province of Manitoba. Safe Software in particular, with assistance from other participants, was able to perform analyses of EO data and develop indicators for flood extents and impacts in the region.

The Canada scenario focused on flooding and flood impacts in the Red River Basin, an area prone to frequent flooding due to typically rapid spring melting, relatively flat slope, northward flow, frozen ground and frequent ice jams. The Red River flows north from Northeast South Dakota and West Central Minnesota into Manitoba Canada and eventually out into Hudson Bay. A typical spring thaw occurs from the middle of March across southern portions of the basin and mid or late April across the north. Data from a flooding event in 2011 were used for analysis.

After looking at a range of possible indicators, the goal for the recipe developed for the Red River Scenario was to generate an indicator of best transportation routes to avoid flood water, along with recipes for the supporting data value chain. The value chain starts with observations of flooding using both river gauge measurements and satellite EO observations. These datasets are combined with mathematical algorithms and a Digital Elevation Model (DEM) which describes the height of the land, to produce two different approaches for the creation of a flooded area ARD.

7

LESSONS LEARNED

LESSONS LEARNED

7.1. Using GeoPackage Security

> **IMPORTANT**
>
> *Secure GeoPackage offers new distribution channels to enable trusted data sharing for disaster management across different stakeholders.*

The use of Secure GeoPackage allows a fine grained encryption of data items to be stored in a GeoPackage. The encryption of the data items can be controlled using policies that fit the disaster management requirement but also reflects the sensitivity level of the data. For example, data items could be encrypted with different keys based on their sensitivity or classification level, their location (relative to the location of the disaster), privacy implications, etc. As an extreme, all data items could be encrypted with the same key, only one data item could be encrypted or even each data item in a Secure GeoPackage could be encrypted with a different key. As access to each key can be controlled via the Key Management System, great flexibility is provided which users are able to decrypt the data items. The Secure GeoPackage — as a secure data container — can then be distributed by any means including USB-sticks, shared network drive, mobile devices or even streamed

Establishing the data exchange via Secure GeoPackage (however) requires to deploy additional services.

- **Secure GeoPackage Service** is a required service to produce GeoPackage(s) with encrypted content.

- **Key Management System** is a required service that controls access to the (de)encryption keys. Enforcing access to the keys is based on user authentication. To ensure seamless access management, all possible users must be identified by system-wide unique identifiers.

- **Authorization Server** is a required service that provides unique identifiers to users.

> **IMPORTANT**
>
> *Authentication is key!*

Despite all the other challenges when creating and using Secure GeoPackage(s), as outlined in section GeoPackage Security, strong and trustworthy authentication is a must to success. As illustrated in the architecture figure, the guarantee of (globally) unique user identifiers is paramount to ensure access management when creating Secure GeoPackages and setting

conditions for key access. For this project, the use of an AaaS (Authentication as a Service) is used that supports federated identity management but guarantees unique user identifiers.

> **IMPORTANT**
>
> *Key Management is key!*

Any security solution that is based on encryption introduces quite some challenges that need to be considered in architecture, implementation and deployment. But, at the end of the day, it is ultimate to keep the keys save! Leak of the keys jeopardizes the entire solution. Therefore, the Key Management System shall provide a very simplistic, non-chatty and well tested API.

7.2. GeoCollaborate

7.2.1. Challenges and lessons learned

7.2.1.1. Challenge 1. Identifying consistent web services

While the pilot did develop data layers that were served via OGC compliant web services, those services were not always working since they were being tested and updated fairly frequently. This provided occasional challenges when GeoCollaborate was being demonstrated with DP21 data layers. When the service is not working, the data layers will not be accessed or shared. Perhaps future pilots could provide a standard location to host services more consistently so demonstrations could be provided at any time and the up-time percentage of those servers could be close to 100%. It is to be expected that web services will be updated frequently during a Disasters Pilot. This is where innovation happens and is much more experimental than operational.

7.2.1.2. Challenge 2. More work needs to be done to assess the value of the GeoPackage data format so it is friendly for collaboration.

While GeoPackage provides a secure way to share data and package up certain sensitive data for sharing, there are also other 'secure' environments evolving that could provide a rapid 'cloaked' capability to initiate collaboration sessions that are invisible to any non-invited collaboration participant. These cloaked sessions could be very attractive to organizations in both the civil, defense and intelligence sectors and use cases could be developed that look at the value of creating a 'cloaked' environment vs. generating GeoPackages that keep data and/or pieces of sensitive data secure for sharing. When rapid collaboration sessions need to be initiated, establishing an entirely 'cloaked' environment may be mush easier than making sure all data providers are using the GeoPackage data format.

7.2.1.3. Challenge 3. Getting sponsors & stakeholders to participate in technology demonstrations

Stakeholder meetings are valuable so participants can report on progress made during the OGC pilot. It is much more valuable when the stakeholders/sponsors actually participate actively by asking questions and inviting innovation that can directly impact their goals. When presentations are given and no feedback is delivered by the stakeholders and/or sponsors, the ability to move forward is limited by simply executing on the funded proposal. More progress could be shown with heightened interactivity.

7.2.1.4. Challenge 4. Engagement of end users during pilot projects

If future pilot project can work to engage users ahead of time and keep them engaged through the pilot, we believe that the implementation and success rates of OGC pilots will increase. It would be good to see what the end goal is of the sponsors so participants can deliver on those goals or at least make progress toward those goals. Engaging the end users can go a long way in making this successful for everyone. Sponsors would like solutions and participants would like sponsors to use/purchase or license their technologies. That is a win/win/win for everyone.

7.2.1.5. Challenge 5. Telling a comprehensive story about how all elements of DP21 (and future Pilots) serve the sponsors

There were so many excellent activities going on within the DP21 project it was tough to communicate how those pieces fit together to deliver a pathway forward for sponsors and users. Identifying gaps and areas of future activities will power future Disaster Pilots that will benefit sponsors, users and members of OGC. It would also be good to understand how the DP21 and future pilots fit into the sponsor's plans and objectives. We are confident that this will evolve as future pilots get funded and executed.

7.2.2. Lessons learned

This was the very first project where StormCenter participated as a funded OGC participant. There was a lot to learn including how to navigate and publish reports and updates using the web portal and Wiki while also learning the basics of Asciidoc publishing. The weekly meetings were good to stay engaged and provide updates. It might have been beneficial to place some of the participants in direct communication with the sponsors or have sponsor-specific meetings to discuss progress and whether or not they had questions about the progress of the project or of any individual activity being funded. Some sponsors were hard to connect with and others seemed to keep 'arms-length' distance which may prevent the open exchange of information that could guide progress.

The development of the 'Guides' is an excellent step forward to make data producers and data users better understand how to think about producing and using trusted data sets to drive decision making.

7.3. Routing and Logistics

- In order to estimate the roads affected by a flooding at a specific moment, apart from the flood extent, the flood depth is a critical factor. However, in some scenarios it might be difficult to gain access to satellite imagery of sufficient accuracy, which was the case in the Rimac River scenario. Other sources of data can be considered to estimate the flood depth as an alternative or a complement to satellite imagery, such as river and flood gauges, inputs from public and first responders (such as the one proposed by GISMO) or modelings based on terrain topography.

- As demonstrated in this pilot, transforming flooding ARD (flood contours and depth) into DRI (roads that should be closed) needs a series of parameters. The value of these parameters depend on the local context and should be adaptable. For example, the maximum flood depth for emergency vehicles will be higher in municipalities that have invested in equipment that is capable to wade deep waters. The maximum flood depth for public vehicles, which depends on the typical ground clearance of cars, will be higher in rural America than in urban Europe.

- The case of flooding disaster prevention in roads was considered, particularly for the scenario of flash flooding in roads without adequate drainage that can pose a mortal risk for drivers. However, it does not seem to be feasible using satellite imagery and it would require visual inspection on the field by expert personnel.

8

FUTURE WORK

---

8 FUTURE WORK

8.1. Key Management System API Standardization

A Key Management System must be secure but also interoperable. Its API must be simple but flexible. The implementation from Secure Dimensions, based on NIST 800-57, was proven to do a "pretty good" job in Disaster Pilot '21. However, the API is not standardized. In order to ensure the productive use of Secure GeoPackage, the standardization of the KMS API is paramount.

8.2. Distributed User Identifiers

For any disaster management, where users from different stakeholders use different authentication systems, the controlled and secure exchange of relevant data is almost impossible. Why? Because the acceptance of user identifiers and the management of user privileges and access rights in a distributed / federated system is very challenging, if not impossible.

As the modification of existing systems is typically not possible, new ideas for distributed user authentication must be explored. First solutions in distributed authentication based on distributed ledgers (based on a Blockchain) are in production since Bitcoin.

In a future pilot, it should be explored how the use of Distributed-Ledgers based authentication can improve the management of user access rights for ensuring a secure data exchange in disaster management. The following publication could be used for guidance: Blockchain Identity Management: The Definitive Guide (2021 Update).

8.3. Using GeoCollaborate

GeoCollaborate can play a significant role in connecting OGC pilot data services, including ARD and DRI, into decision making environments. This could introduce decision makers to the significance of web-based data services, OGC data standards and the burden it removes from decision makers when data is assigned specific Operational Readiness Levels (ORLs). ORLs have been developed by the ESIP Federation and the All-Hazards Consortium (AHC) to provide a level of confidence in data quality and trust to accelerate situational awareness and decision making by non-science experts. While much attention is given to the systems, algorithms and processes to produce ARD and DRI, not enough attention is given to engaging the end users and decision makers so the data can be put to use effectively and efficiently.

8.4. EO Value Chains

During the activities performed in this pilot, Terradue were able to define the EO value stream flow to support decision makers to use Earth Observation. The EO value stream flow derives the information contained in EO products and acts as an interface between the technical characteristics of the satellite and the specific needs of the decision maker. Taking advantage of the Earth Observation Exploitation Platforms and the available Cloud technologies, the flow goes from Acquisition to Decision. It defines steps for product acquisition, ingestion, calibration and the actual production of Analysis Ready Data or even more importantly Decision Ready Information from multiple satellite missions.

In the activities performed we identify the Technical Components necessary for the product ingestion and calibration steps of the EO value stream flow. Further work is necessary to identify and demonstrate the full flow reaching the production of the Decision Ready Information. In particular, further activity is necessary to extend the catalog (e.g., STAC) with the necessary information model ready to support the whole Disaster Reduction and Response cycle. Conversely one key point that was identified is that even if reproducibility is most surely provided by the Application Package it is necessary to provide exact lineage that links the information produced with the original EO products.

From the Disaster Response perspective we also identified the need for a process to identify and generate reference datasets during a disaster. For example, while responding to a flood event we are able to identify the necessary dataset and processing to obtain the exact flood extent but not so straightforward to provide the community with the reference situation previously or not under the influence of the event that they are addressing.

8.5. Disaster Routing

- The integration among different components in the data flow has been tested at the level of data format, but tests were not complete at the level of live data, APIs and processes. In the case of flood contours processing and road closure management, Skymantics Road Risk Estimator could complete the task for the Red River scenario in a matter of seconds, estimated between 10 and 30 depending on the flooding moment. This performance is considered acceptable based on the low frequency of ARD updates and proves the technical feasibility of the concept. In any case, more integration and testing would be necessary at the service level to fully demonstrate the real-time flood of information.

- The potential to use routing and travel time calculations for disasters was tested successfully in two health related applications, one for supply routes and the other to estimate distance to care for entire districts. There is still a lot of potential to be uncovered combining routing and distance calculation capabilities in specific domain applications, particularly in planning and relief. For example, finding the optimal location for shelters, supply depots or temporary hospitals based on minimal travel distance for targeted vulnerable population.

- The case of landslides poses a challenge to the prediction and terrain analysis tools. However, from the point of view of road management it was considered a simplification of the flooding scenario, solved by just blocking the roads affected after the hazard. The case of wild fires can be more complex and of great interest for future work, with several different ARDs to be considered for a proper road management, starting with roads with visibility affected by smoke, to the decision of closing a road based on the distance to fire and the expected evolution of the wildfire.

8.6. Voice Survey

During the Pilot, GISMO were able to engage a geo-located, multi-lingual survey platform, which will be accessible here for use: https://app.ellaxyz.com/.

There are a number of other solutions to build out in the same vein as Ella, which all revolve around the singular idea of technology removing barriers between community and decision maker.

SUMMARY AND CONCLUSIONS

9 SUMMARY AND CONCLUSIONS

Disaster Pilot 2021 was a rousing success in terms of engaging both technical participants and regional / thematic stakeholders. It advanced the concepts of ARD, DRI, Cloud applications-to-the-data, collaborative DRI recipe development, linked data Web search optimization, multilingual voice surveys of impacted populations, and data-secure GeoPackage sharing. Most of the participants would also assert that the Pilot was only able to scratch the surface of each of these capabilities. Most have provided lists, in some cases quite extensive, of next steps and future work to extend the work they accomplished in DP21, to more regions, other hazards, more complications, different indicators, and more operational readiness in terms of scale or sustainability. Most of all, the participants would agree that work remains to be done to get these data sharing and actionable guidance capabilities into the hands of those who really will need them.

Three additional points should be made.

1. Timeliness: while a number of ARD→DRI recipes were successfully executed and some near real-time data were accessed, the Pilot was generally unable to demonstrate acquisition and processing of EO data in a timely enough fashion to meet most conceivable disaster awareness timelines. While cloud computing might provide the necessary processing throughput, without availability of timely input data, scalability is for nought.

2. Preparation: In many / most cases, data sources which were not available at the start of the Pilot were still not available by the end. This is true of actual disasters as well. Data sources and other resources, as well as expertise and sharing agreements, clearly need to be in place and practiced for there to be any hope of successful throughput in an actual emergency

3. Response: despite serious outreach efforts, it is clear that most first responders are in fact busy responding, and do not have the time or access to engage in forward-looking but to some degree speculative activities such as OGC Pilots. This makes it difficult to figure out what would work best for them in terms of either access and applications, or useful indicator information. An important challenge for follow-on activities is to identify or create a channel for responder engagement. One possibility is to structure such engagement as a form of responder training. To be fair and responsible, such training would need to convey some skill or knowledge which could be immediately useful in a first responders daily responsibilities.

A

ANNEX A (INFORMATIVE) GEOPACKAGE ENCRYPTION EXTENSIONS

ANNEX A
(INFORMATIVE)
GEOPACKAGE ENCRYPTION EXTENSIONS

NOTEThe definition of the extensions are almost identical and therefore contain duplicated information. But the editor has chosen this documentation approach to make it clear that these extensions are actually independent from each other and that each will be packaged into a separate OGC document when being submitted to the GeoPackage SWG for starting the standardization effort.

The GeoPackage Encryption Extensions are the result from work within the OGC Disaster Pilot '21: The first extension defines SQLite schema rules for storing encrypted Features and the second extension for storing encrypted Tiles. Both extensions define how encrypted data (features or tiles) and the associated decryption key(s) as well as related metadata information can be stored.

The client implementation supporting the decryption of features and tiles was implemented as QGIS plugins in Python. The GeoPackage Encryption Service (GES) is identified as deliverable D111-3 within the Disaster Pilot. The prototype implementation of this service (GES) is realized as a plugin for GeoServer, version 2.20. The plugin extends the OGC Web Map Service, version 1.1.0 interface to produce GeoPackage files including encrypted Tiles in PNG format and the OGC Web Feature Service, version 2.0.0 interface to produce GeoPackage files including encrypted Features with source encoding in GeoJSON.

The plugin was implemented by Secure Dimensions and is available as Open Source from Github: GeoServer Encrypted GeoPackage Plugin.

An OpenAPI description of the GES was developed to illustrate the extended use of OGC WMS and WFS provided by Geoserver, v2.20. For the Disaster Pilot, a simple demonstration that produces a GeoPackage with encrypted content for OSM places US is available for the GES OpenAPI. The underlying GeoServer deployment is available here.

A.1. GeoPackage Encryption Extension for Features

> **WARNING**
>
> *This extension is an OGC Disaster Pilot '21 result and may change radically.*

A.1.1. Extension Title

GeoPackage Encryption Extension for Features

A.1.2. Introduction

This extension provides a storage scheme that allows to store Features, originally encoded in GeoJSON, as a SQLite BLOB. Each row in the features table(s) contains the encrypted Feature data and references to the metadata of the Data Encryption Key (DEK) via a foreign key to the `gpkg_ext_keys` table. The DEK data is encoded as a JSON Web Token ([RFC 7519](#)) or JSON Web Encryption ([RFC 7516](#)). As such, the JWT is digitally signed by the entity that created the key's metadata to ensure integrity and authenticity. The payload of the JWT contains information about the key and in particular a `kurl` that provides the link to fetch the DEK from the corresponding Key Management System (KMS). A JWE encoded information indicates that the actual DEK is encrypted in the JWE's payload using the JWK format.

A.1.3. Extension Author

Secure Dimension GmbH, in collaboration with the participants of the OGC Disaster Pilot 2021.

A.1.4. Extension Name or Template

`sd_encrypted_features` (will become `gpkg_encrypted_features` if adopted by the OGC)

A.1.5. Extension Type

New requirement dependent on [GeoPackage Core (Clause 1)](#) and [GeoPackage Schema Extension](#).

A.1.6. Applicability

This extension allows to store Simple Features, originally encoded in GeoJSON, as an encrypted BLOB. This extension does not constraint how the data is encrypted. It is outside the scope of this specification how the cipher key information is protected and securely communicated to the client application that is supposed to decrypt the data. Information about the encryption method (algorithm) and key length is available from the metadata of the DEK that can be obtained from the KMS in case a key reference is stored in the `gpkg_ext_keys` table or directly from the row in the `gpkg_ext_keys` table in case a JWE format is used.

NOTEFor the OGC Disaster Pilot 2021, a working prototype was implemented that is described in more detail in the Disaster Pilot Summary ER.

A.1.7. Scope

read-write

A.1.8. Specification

The following sub-sections define how to store the encrypted content and the key information in a GeoPackage.

A.1.8.1. Table `gpkg_extensions`

To use this extension, add the following rows to this table.

Table A.1 — gpkg_extensions Table Rows

T COLUMN N NAME	EXTENSIC NAME	DEFINITION	SCOP
g e null k	sd_ encrypte features	https://www.ogc.org/per/021-064.html#sd_encrypted_features	read-write
< d e null f t	sd_ encrypte features	https://www.ogc.org/per/021-064.html#sd_encrypted_features	read-write

NOTE The values in the `definition` column SHOULD refer in some human-readable way to this extension specification. If the extension is adopted by the OGC, it will gain the "gpkg_" prefix (replacing "sd_") and get a different definition permalink.

A.1.8.1.1. Table `gpkg_data_columns`

This table contains the information about the data type for the columns in the `<user-defined encrypted features table>` and `gpkg_ext_keys` table.

To use this extension, add the following rows to this table.

Row 1: Definition of the `data` column from the `<user-defined encrypted features table>` table

- `table_name` has the value `<user-defined encrypted features table>`

- `column_name data`

- name has the value `<user-defined encrypted features table>-data`

- title has value `Encrypted Feature Data`

- description description, e.g. `The encrypted data of the feature`

- mime_type has value `application/octet-stream`

- constraint_name has the value `null`

Row 2: Definition of the `data` column from the `gpkg_ext_keys` table

- table_name has the value `gpkg_ext_keys`

- column_name data

- name has the value `sd_encrypted_features-keys`

- title has value `DEK metadata`

- description description, e.g. `The Data Encryption Key information represented as JWT or JWE`

- mime_type has value `application/jose` to represent JWS, JWT or JWE encoding of the key's metadata

- constraint_name has the value `null`

NOTE No explicit media type exists for JWS or JWE encoded data. IANA media type "application/jose" must be used and from counting the dots, an application can defer whether the structure refers to JWS / JWT (2 dots) or JWE (4 dots). Some applications use the value "JWE" and "JWS" or "JWT" to determine between JOSE encodings. For this extension, the IANA media type "application/jose" is used.

A.1.8.2. New Table Definitions

Following are definitions of the tables for this extension.

As with other GeoPackage tables, this extension takes no position on how either of these tables are to be used by a client.

A.1.8.2.1. Table `<user-defined encrypted features table>`

NOTE 1 The place holder `<user-defined encrypted features table>` refers to the name of the table that actually holds the encrypted feature information. This name varies but is identical for a GeoPackage instance.

This table contains encrypted features. The identifier is a string, uniquely representing the feature stored in this table. The identifier may be an obfuscation of the actual feature type in cases where naming the table after the feature type might disclose sensitive information.

NOTE 2   For the Disaster Pilot '21 implementation, the table name corresponds to the `featureType` parameter of the WFS request and the identifier (fid) is the id of the actual feature that is encrypted.

The columns of this table are:

- `id` (type INTEGER) is the primary key

- `fid` (type TEXT) is the value of the original feature id

- `the_geom` (type BLOB) is the bounding box of the encrypted feature or null in case it is withheld for sensitivity reasons or not available

- `data` (type BLOB) is the encrypted feature

- `kid` (type TEXT) is the foreign key to the entry in the `gpkg_ext_keys` table

A.1.8.2.2. Table `gpkg_ext_keys`

This table contains the Data Encryption Key or its metadata. The primary key is used as foreign key by the `key_id` column from the `<user-defined encrypted features table>`.

The content of the `data` column contains either a JWT or a JWE. Typically, the format is either JWT containing metadata when managing the DEK at some other entity or a JWE when the DEK is encrypted inline.

The columns of this table are:

- `id` (type TEXT) is the primary key

- `data` (type TEXT) is the JWK representation either as JWT or JWE

A.2. GeoPackage Encryption Extension for Tiles

> **WARNING**
>
> *This extension is an OGC Disaster Pilot '21 result and may change radically.*

A.2.1. Extension Title

GeoPackage Encryption Extension for Tiles

A.2.2. Introduction

This extension provides a storage scheme that allows to store Tiles, originally encoded in PNG, as a SQLite BLOB. Each row in the tiles table(s) contains the encrypted tiles data and references to the metadata of the Data Encryption Key (DEK) via a foreign key to the `gpkg_ext_keys` table. The DEK data is encoded as a JSON Web Token ([RFC 7519](#)) or JSON Web Encryption ([RFC 7516](#)). As such, the JWT is digitally signed by the entity that created the metadata to ensure integrity and authenticity. The payload of the JWT contains information about the key and in particular a `kurl` that provides the link to fetch the DEK from the corresponding Key Management System (KMS). A JWE encoded information indicates that the actual DEK is encrypted in the JWE's payload using the JWK format.

A.2.3. Extension Author

Secure Dimension GmbH, in collaboration with the participants of the OGC Disaster Pilot 2021.

A.2.4. Extension Name or Template

`sd_encrypted_tiles` (will become `gpkg_encrypted_tiles` if adopted by the OGC)

A.2.5. Extension Type

New requirement dependent on [GeoPackage Tiles Option (Clause 2.2)](#) and [GeoPackage Schema Extension](#).

A.2.6. Applicability

This extension allows to store tiled data, originally encoded in PNG, as an encrypted BLOB. This extension does not constraint how the data is encrypted. It is outside the scope of this specification how the cipher key information is protected and securely communicated to the client application that is supposed to decrypt the data. Information about the encryption method (algorithm) and key length is available from the metadata of the DEK that can be obtained from the KMS in case a key reference is stored in the `gpkg_ext_keys` table or directly from the row in the `gpkg_ext_keys` table in case a JWE format is used.

NOTE For the OGC Disaster Pilot 2021, a working prototype was implemented that is described in more detail in the Disaster Pilot Summary ER.

A.2.7. Scope

read-write

A.2.8. Specification

A.2.8.1. Table `gpkg_extensions`

To use this extension, add the following rows to this table.

Table A.2 — gpkg_extensions Table Rows

TAL COLUMN NA NAME	EXTENSION NAME	DEFINITION	SCOPI
gpl ext null key	sd_ encrypted tiles	https://www.ogc.org/per/021-064.html#sd_encrypted_tiles	read- write
\<us def enc null til tal	sd_ encrypted tiles	https://www.ogc.org/per/021-064.html#sd_encrypted_tiles	read- write

NOTE The values in the `definition` column SHOULD refer in some human-readable way to this extension specification. If the extension is adopted by the OGC, it will gain the "gpkg_" prefix (replacing "sd_") and get a different definition permalink.

A.2.8.2. Table `gpkg_data_columns`

This table contains the information about the data type for the columns in the `sd_encrypted_tiles-<user-defined encrypted tiles table>` table.

To use this extension, add the following rows to this table.

Row 1: Definition of the `data` column from the `<user-defined encrypted tiles table>` table

- `table_name` has the value `<user-defined encrypted tiles table>`

- `column_name` data

- `name` has the value `<user-defined encrypted tiles table>-data`

- `title` has value `Encrypted Tile Data`

- description description, e.g. `The encrypted data of the tile`

- `mime_type` has value `application/octet-stream`

- `constraint_name` has the value `null`

Row 2: Definition of the `data` column from the `gpkg_ext_keys` table

- `table_name` has the value `gpkg_ext_keys`

- `column_name` data

- name has the value `sd_encrypted_tiles-keys`

- title has value `DEK metadata`

- description description, e.g. `The Data Encryption Key information represented as JWT or JWE`

- `mime_type` has value `application/jose` to represent JWS, JWT or JWE encoding of the key's metadata

- `constraint_name` has the value `null`

NOTENo explicit media type exists for JWS or JWE encoded data. IANA media type "application/jose" must be used and from counting the dots, an application can defer whether the structure refers to JWS / JWT (2 dots) or JWE (4 dots). Some applications use the value "JWE" and "JWS" or "JWT" to determine between JOSE encodings. For this extension, the IANA media type "application/jose" is used.

A.2.8.3. New Table Definitions

Following are definitions of the tables for this extension. As with other GeoPackage tables, this extension takes no position on how either of these tables are to be used by a client.

A.2.8.3.1. Table `<user-defined encrypted tiles table>`

NOTE 1 The place holder `<user-defined encrypted tiles table>` refers to the name of the table that actually holds the encrypted tile. This name varies but is identical for a GeoPackage instance.

This table contains encrypted tiles. The identifier is a string, uniquely representing the tile layer stored in this table. The identifier may be an obfuscation of the actual tile layer in cases where naming the table after the tile layer might disclose sensitive information.

NOTE 2 For the Disaster Pilot '21 implementation, the table name corresponds to the `layer` parameter of the WMS request.

The columns of this table are:

- `id` (type INTEGER) is a primary key

- `zoom_level` (type INTEGER) as specified by GeoPackage

- `tile_column` (type INTEGER) as specified by GeoPackage

- `tile_row` (type INTEGER) as specified by GeoPackage

- `tile_data` (type BLOB) is the encrypted tile using symmetric key encryption with the key described in `dek_info` column.

- `kid` (type STRING) is the foreign key to the entry in the `gpkg_ext_keys` table

A.2.8.3.2. Table `gpkg_ext_keys`

This table contains the Data Encryption Key or its metadata. The primary key is used as foreign key by the `id` column from the `<user-defined encrypted tiles table>`.

The content of the `data` column contains either a JWT or a JWE. Typically, the format is either JWT containing metadata when managing the DEK at some other entity or a JWE when the DEK is encrypted inline.

The columns of this table are:

- `id` (type TEXT) is the primary key

- `data` (type TEXT) is the JWK representation either as JWT or JWE

B

ANNEX B (INFORMATIVE) REVISION HISTORY

———

B ANNEX B (INFORMATIVE) REVISION HISTORY

DATE	RELEASE	AUTHOR	PRIMARY CLAUSES MODIFIED	DESCRIPTION
October 01, 2021	0.1	A. Matheus	all	added D111-3
September 21, 2021	0.2	A. Matheus	all	added GeoPackage Encryption Extension
January 12, 2022	0.3	A. Matheus	10-11, lessons-future, geopackage-security	annex-d111-3 moved into section 10.11, geopackage security and related aspects documented in Lessons Learned / Future work
March 15, 2022	0.4	A. Matheus	almost all	editorial aspects for harmonizing contributors sections