DELIVERABLE D2.3

Grant Agreement number : CIP-297300
Project acronym : InGeoCloudS
Project title : INspiredGEOdata CLOUD Services

Funding Scheme : Pilot B

InGeoCloudS Web Services Covering Use Cases

D2.3

Version 1

Reference **D2.3-INGC**

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contract Number :	CIP-297300
Document Title :	InGeoCloudS Web Services Covering Use Cases
Document version :	1
Document status :	Approved
Date :	2013-09-11
WP contributing to the deliverable :	WP2
Availability :	Public
Authors :	Jørgen Tulstrup, Jakob Molander, Pierre Lagarde, Benoit Baurens, Jasna Šinigoj, Martin Podboj, Kiriaki Konstantinidou, Makis Atzemoglou and Kyriakos Kritikos.
Approved by :	InGeoCloudS Steering Committee

Abstract

This document describes the web services that have been developed and implemented on the InGeoCloudS platform with the purpose of supporting the use cases that have been selected in the project. It is described how the requirements specified for the different use cases have been mapped to technical requirements that the InGeoCloudS platform makes available. Furthermore the implementation of the specific web services for the use cases are described including how these services fit into the InGeoCloudS overall architecture. For some of the services their implementation in InGeoCloudS required additions to the lower level API functions.

Keywords List

InGeoCloudS Architecture; InGeoCloudS deployment; InGeoCloudS Data Management; InGeoCloudS Basic Services; Linked-Open Data; Data Publication; Scalability.

DOCUMENT CHANGE LOG

Document Issue.	Date	Reasons for change
Version 1-Draft 1	2013-06-20	Creation of the document
Version 1-FinalDraft	2013-09-09	Following Draft7 after final integration of last contributions for EKBAA's Use Cases
Version 1-Approved	2013-09-11	After last editorial changes and checks

APPLICABLE AND REFERENCE DOCUMENTS (A/R)

A/R and Document Reference	Title
[A1] ICT PSP Grant Agreement N° CIP 297300	InGeoCloudS Grant Agreement and its annex (including the description of work)
[R1] D3.1.1-INGC	Analysis and Monitoring of clouds for geo-data services
[R2] D2.1-INGC V1.1. (re-submitted – 08.05.2013)	Use Cases for InGeoCloudS data and services
[R3] D2.2-INGC	Interface of Web Services and models of data
[R4] D3.2-INGC	Cloud architecture configuration and data access implementation
[R5] CS-INGC-02 and CS-INGC-04	Replication mechanisms and High-availability architecture with Postgresql See http://www.InGeoCloudS.eu/?q=task-32-infrastructure-and-data-access-functionnalities/replication-postgresql
[R6] CS-INGC-08	WebGIS Client Integration: "How to" guide in order to let Data Providers integrate their client-side applications in INGC's default WebGIS Client
[R7] D5.1-INGC	InGeoCloudS Users Documentation
[R8] D6.5.1-INGC v2	InGeoCloudS Pilot Exploitation Strategy

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1. ACRONYMS AND DEFINITIONS

Term	Definition
Amazon AWS	Amazon Web Services, i.e. the amazon cloud computing infrastructure
API	Application Programming Interface
Data Provider	A user willing to contribute to InGeoCloudS with his own data or with a novel service
Data Provider Service	Web-accessible service based on InGeoCloudS system
DOI	Digital Object Identifiers
GDAL	Geospatial Data Abstraction Library
GFS	Gluster File System
INGC	InGeoCloudS
LDAP	Lightweight Directory Access Protocol
LOD	Linked Open Data
N/A	Not Applicable
NFS	Network File System
ODBC	Open Database Connectivity
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
RDF	Resource Description Framework
Registered User	Visitor of the platform that performed the registration process
REST	REpresentational State Transfer
SPARQL	SPARQL Protocol and RDF Query Language
SPARUL	A.k.a. SPARQL/Update, is a declarative data manipulation language extending SPARQL
URIs	Universal Resource Identifiers
WFS	Web Feature Service
WMS	Web Map Service
Workspace	Provides to data providers the possibility to store (and access) own data in a private database of the InGeoCloudS Elastic Database, or in a private folder of the InGeoCloudS Elastic File Server

Table 1.Useful acronyms and Definitions

1.2. OBJECTIVES OF THE DOCUMENT

The purpose of this document is to give potential new providers of data and/or services insight into the work related to implementing such data and services into the InGeoCloudS system (in the following called INGC). This is done by describing relevant functionalities available in the INGC Pilot and by describing the kind of work that has been carried out by the consortium partners in order to implement their Use Cases and how the cloud infrastructure deployed by the project has influenced the development / adaptation of each use case. It furthermore contains a description of a web processing service that is under development for Pilot2. The service is capable of calculating a set of grid points from a dataset of randomly distributed data points using the *Kriging* method.

Please note that what is called “web services” or “service” (in the original title and in the rest of this document) is not meant from the technical sense but more as a list of facilities available on the INGC.

The document is not a *user guide* for the services. Documentation of the different applications and services integrated by data providers is available in the portal itself as online help.

It is neither a *cookbook* explaining how a new data provider should develop new services and implement those in INGC. This kind of guidance is available for reference as “Users Documentation” available on www.ingeoclouds.eu where different types of documentations are elaborated in order to match with the different end-user needs (see [R7] for documentation rationale details).

The document builds on D2.1 v1.1 (description of the use cases) and D2.2 (description of the interfaces of the services implemented in INGC). It describes some of the work done in T2.3 “*Service adaptation, modification and creation*”.

According to the DoW, this task should to a large degree be about porting and modifying existing web services running at the data providers own systems and integrating those in INGC. It has however turned out, that a large part of the functionality of the use cases is new. The work in T2.3 therefore has not so much been focused on trying to reuse existing web services and integrating them in the INGC but more on building new services applicable for use in INGC on the top of the INGC middleware architecture but using the already existing datasets. During that process valuable input for the overall architecture and design of the INGC has been obtained.

1.3. OVERVIEW OF THE DOCUMENT

The document goes through the services that are implemented on INGC for Pilot 1 and 2. Both the general (administrative, general GIS functions, etc.) and the use case specific services are described. For each set of services the mapping from the use case requirements mentioned in D2.1 v1.1 ([R2]) is accounted for (Use case numbers are reminded for convenience). Then, a description of the interfaces of the services follows. This description extends the description mentioned in D2.2 ([R3]). Finally it is described how the services fit into the INGC infrastructure and whether their implementation has had implications for the overall design of the INGC system.

The document finally contains some notes about currently identified limitations and future work.

2. DOCUMENTATION OF SERVICES

The INGC services are divided into three groups:

1. General services for managing and operating the INGC system, general GIS functions, etc.
2. Services that support the different use cases.
3. Processing services.

2.1. GENERAL SERVICES

2.1.1. ADMINISTRATIVE (USER MANAGEMENT) SERVICES

The following administrative functions are provided by the *INGC Management* component and they support the management of InGeoCloudS users. These functions are mainly related to the creation of user accounts, and their management over time, i.e. update of user information and account removal.

2.1.1.1. Description of Functionalities implementation

Function	Description	Documentation
Request Account	Records the request for the creation of a new *data provider/registered user* account. All requests have to be validated (e.g.) by email confirmation. Additionally, data provider requests have to be evaluated and approved by the InGeoCloudS administration.	Account creation is implemented by the /master/provider API as described in http://ingeoclouds-api.isti.cnr.it/platform/resource_Master%20Service.html#path__master_provider.html This invokes other methods from the Elastic Database Layer and the Elastic File System layer to create the Data Provider Workspace.
Confirm Account	Accounts are created after moderation by INGC administration	The INGC Administration module relies on INGC's OpenAM implementation for the actual creation of accounts in the system. Notification to the user about any change in the account status can be configured to happen by email.
Update Account	Updates the information of a given *data provider/registered user*	Account update is implemented by the /master/provider API as described in http://ingeoclouds-api.isti.cnr.it/platform/resource_Master%20Service.html#path__master_provider_-id-.html
Delete Account	Deletes the account of the given *data provider/registered user* and removes his private workspace (if data provider)	Account deletion is implemented by the /master/provider API as described in http://ingeoclouds-api.isti.cnr.it/platform/resource_Master%20Service.html#path__master_provider_-id-.html This invokes other methods from the Elastic Database Layer and the Elastic File

		System layer to delete the Data Provider Workspace.
Retrieve Account	Retrieve the information about a given *data provider/registered user* stored in the system	This is implemented by the /master/provider API as described in http://ingeoclouds-api.isti.cnr.it/platform/resource_Master%20Service.html#path__master_provider_-id-.html
Register Notifications	Enable the user to receive notifications.	This service is not implemented yet: some generalisation of the function is under discussion and shall be introduced to the platform by end 2013.
Update Notifications	Updates the notifications criteria	This service is not implemented yet: some generalisation of the function is under discussion and shall be introduced to the platform by end 2013.
Remove Notifications	Disable the notifications for the given user	This service is not implemented yet: some generalisation of the function is under discussion and shall be introduced to the platform by end 2013.

2.1.1.2. Integration within the INGC infrastructure

The user management service is integrated into the INGC Management component and it can be considered as one of the lowest levels of the architecture. Its functionalities are provided using an LDAP server, which deliver access and management of the user accounts over an Internet protocol network, so that the users' credentials are shared between all the services of the INGC infrastructure. On top of LDAP it was deployed a specific user management service that let an INGC administrator create/modify/delete provider accounts into the LDAP server. When a new provider account is created, the service will be involved also in the creation of a private space into the distributed file system and into the database so that the provider's data can be moved into the platform preserving data isolation with other providers.

The integration in the INGC infrastructure has been done in pilot 1 and improved in pilot 2.

The high level services provided by this components are available only to an INGC administrator and all the machines being part of the platform use the low level authentication centralized services.

2.1.1.3. Concluding Remarks On InGeoCloudS Integration Work

The user management service currently answers the followings use cases: create geo-publishing account, publishing geo-dataset and manage common geo-services. It was deployed without major difficulties in the cloud infrastructure, using standard and open source APIs and tools, such as OpenLDAP. In general, the service is involved with all the operations regarding the creation and the administration of the provider accounts, and it can be used for providing access for any possible use case / application.

2.1.2. GENERAL GIS SERVICES (GIS CLIENT FUNCTIONS FOR DATA VISUALIZATION)

INGC features a WebGIS Client tool also called the "*default WebGIS Client*" as it aims at providing a simple yet fully-functional tool for data visualisation on maps. It provides visualisation of geolocalized points and geometries on a default base (OpenStreetMap).It supports WMS-defined layers display and selection and enables - without coding - the integration of additional GUI elements like forms and HTML displays in side panels.

- It is a client-side application that is fully integrated in the Portal and provides a web mapping service based on usage of a background map and a WMS layer over it. It includes regular end-user interactive tools (zoom, pan…) on the produced maps. The tool is primarily targeted at maps consultation**.**
- Note that the DataPublication module includes a "*Map service*" (see section 5.3 in [R2]), primarily targeted at **maps design** ("data portrayal" management)**,** allowing for the creation and customisation – without coding - of a Web GUI displaying a set of layers (OGC services) as a final map.

2.1.2.1. Mapping From Use Case requirements To Technical Requirements

<table>
<tr><th>D2.1 Ref.</th><th>Description of user requirement</th><th>Evaluation</th></tr>
<tr><td rowspan="3">Table3, #5
And
Table 5, #3</td><td>Web-application should contain a map with the following functionalities:
1. Zoom in and out.
2. Pan.
3. Display topographical background map.
4. Display other background maps of user's choice including Corine Land Cover and OneGeologyEurope surface geology.
5. Enable user to select and display other maps available as WMS/WFS</td><td>All functions are available.</td></tr>
<tr><td colspan="2">Resulting technical requirements and technical choices</td></tr>
<tr><td colspan="2">Both tools use OpenLayers libraries and allow for displaying WMS generated maps on a default background. Navigation and zooming and geometry drawings are also possible and in the case of the DataPublication WebGIS Client, fine customization of data portrayal is possible.
Geoext andExtJS-based developments propose project templates and allow for easy integration of client-side software. OpenLayers APIs also offer state-of-the-art mechanisms for interactions with maps and data (selection, format transformation…).</td></tr>
</table>

2.1.2.2. Description of Functionalities implementation

Function	Description	Documentation
Zoom and display all active layers	Draw rectangle on map when user drags with left-mouse button pressed, Redraw map zoomed to selected area.	Buttons are provided in the menu bar above the maps. Side-panels also allow for selection of layers and for choosing their visibility on the map. OpenLayers API is used behind the scene.
Zoom-out and display all	Draw rectangle on map	Ditto

active layers	when user drags with left-mouse button pressed, Redraw map zoomed-out to area reflecting size of area the user drew (small area drawn = high zoom-out factor)	
Pan and display all active layers	Move origin of map when user drags with left-mouse button pressed, Redraw map with new centre.	Ditto
Select rectangular area	Draw rectangle on map when user drags with left-mouse button pressed	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Display background map	Redraw map which has been selected by user	Ditto
Select map projection	Redraw map projected according to specified map projection	Ditto
Highlight rectangular area	Draw rectangle in area	Ditto
Add list to dropdown box	Add entries to a dropdown box on user interface	Integration and customisation facilities are notably described in a dedicated technical note ([R6]) and a reflected in users documentation
Display the legend	Show the legend of each layer	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Display the overview map	Show the overview of the map in a border of the map	
Display coordinates	Show the X/Y (or long/lat) of the tracker	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Go To…	Allow to zoom in function of a textual criteria	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Print the map	Allow to print and export as image a map	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Request objects in the layers	Allow to request all active layers after selecting an area on the map. Provide all attribute information in the data (DB, FS)	The function is directly available in the tool and relies on OpenLayers API calls and configuration
Fetch list	Fetch a list of values for	

	e.g. a code list from Elastic DBMS	
Fetch background map	Fetch a map layer from Elastic File Server or Elastic DBMS	Background maps are made available in the infrastructure.
Stream data	Stream a collection of objects to the user	Depends on the UseCase application facilities.

2.1.2.3. Integration within the INGC Infrastructure

The WebGIS Client is integrated as an application plugin of the in InGeoCloudS Portal which is based on the the open source Sitools2 framework[1] (see D3.2, section 4.8).

Configuration is made through the INGC administration interface. It also provides a base code structure for organizing specific application domain client code (javascript/extJS) that can be integrated in the proposed panels. Each data provider can thus easily integrate its client-side code (e.g. search/selection forms on its datasets) around the map client display. More technical details are available in a dedicated technical note ([R6]) or in the reference documentation[2].

2.1.2.4. Concluding Remarks On InGeoCloudS Integration Work

The WebGISClient available in Pilot1 provided a simple map consultation tool. It allowed developing a new application developed from GEUS and access to maps and data for the landslides susceptibility use case and integrating without major issues.
Depending on data providers' demands, the WebGISClient module can be proposed with more default base maps and layers, with the integration of additional configuration facilities for the panels, other GUI items, and maps manipulation tools (e.g. new projections).

2.1.3. GENERAL DATA IMPORT SERVICES

These services are part of the Data Publication component as Level 1 services (see section 5.3 in [R2]).

The objective is to provide secured and efficient facilities for directly "pushing" datasets into the infrastructure. Datasets are basically available as relational database contents (that can be incrementally synchronized between providers' infrastructure and INGC's infrastructure, as files and/or as triples (following encoding methods of the Semantic Web).

The Data Import Service, through the use of the Data Import API, allows data providers to import data from some external repository (typically the infrastructure where this data is generated and also stored) into their dedicated workspace inside the InGeoCloudS platform.

2.1.3.1. Mapping from Use Case Requirements to Technical Requirements

D2.1 Ref.	Description of user requirement	Evaluation

[1] http://sitools2.sourceforge.net/
[2] http://www.ingeoclouds.eu/?q=wiki/application-providers-documentation

D2.1 Ref.	Description of user requirement	Evaluation
Table 3, #1 and Table 4, #1 and #2	Pushing of raw data and the execution of the calculation of Susceptibility map of triggering landslides due to rainfall forecast And Automatic/On-demand push of data and calculation of a shakemap in almost real-time for new earthquake events	All functions are available.
	Resulting technical requirements and technical choices	
	After authentication, the Data Import API must provide virtual private workspace to the data provider in order to allow him to push data in the form of data files, database dumps. The FTPS method is supported since Pilot1 and will be complemented with HTTPS in Pilot2. On the server side, the GlusterFS implementation offers secured workspaces. ElasticFS additionally proposes various management methods, default files arborescence, SFTP, a FTP server renowned for its reliability and security is deployed.	
D2.1 Ref.	Description of user requirement	Evaluation
Table 5, #1 and #2	Replication from local database to INGC system must fulfil the following requirements: 3. Be automatic 4. Only include data changed since previous replication Be completed within 2 hours And Data in INGC system should not be more than 1 week out-of-date compared to original data in local database and must be clearly marked with last date of replication from local database.	Replication is available in less than 2 hours
	Resulting technical requirements and technical choices	
	The API harvesting service provides cron-based automatic execution of scripts in the private virtual workspace of each data provider that authenticated to the API. The chosen mechanism is kept relatively simple and allows re-using classical scripting mechanisms that do not impose any particularity of a cloud-infrastructure. Tasks execution times are logged by INGC and status of tasks can be queried at API level (with possible error code if relevant). Please note that based on the indirect import mode in the Linked Data Management API, once a data provider has defined the mapping from his/her relational data to the LD to be stored, any modification to the relational data is immediately and correspondingly performed also on the respective LD.	

2.1.3.2. Description of Functionalities implementation

The Data Import Service currently provides in Pilot2 the following services:

- **/data-import/harvests**: this service is in charge of harvesting data located in data providers' infrastructures and importing these data into the InGeoCloudS platform. Data providers use this service to register a harvesting mechanism that will be triggered at a defined frequency.
- **/data-import/fs/mediaFs**: this method allows data provider to request a temporary access to their dedicated workspace located on file system. Data Provider specifies the transfer protocol to use in a parameter of the REST web service.

Beside "traditional" files and relational database formats, Pilot2 also offers data integration in a triple-store with GSOM (see [R3]) as the core model as part of the *Data Integration&Linking* component.
Two types of Data imports are possible:

- Direct mode: direct importing of data set information through Linked Data (LD) specifications (produced, e.g., using so-call RDFizers): the user is responsible for generating the LD specification (i.e., set of triples) and integrating them directly in the INGC's RDF triple-store (i.e, Virtuoso).
- Indirect mode: Mapping from a relational database encoding is specified thanks to a set of mapping rules expressed in the W3C's standardized R2RML language. The actual RDF triples creation is ensured automatically by the system and subsequent updates on the relational database are also reflected in the triple-store. Alternatively, in the case where the provider's database is in XML, another function has been realized which enables the exploitation of an XSLT specification (mapping from the provider's model to GSOM) that is used for producing the respective RDF triples (in the form of a LD specification) by applying them on each database XML file.

The following methods of the API facilitate the inclusion of data in INGC's triple-store along one or the other method:

- **/linkeddata/ld/ldimport**: This is the direct method for importing LD (corresponding to their existing data sets) for data providers. The LD specification is either provided inline in the request or a URL is provided from which it can be downloaded. The method call can be blocking or not. In the second case, a specific object is returned which contains an import (call) ID and a (import) status message which can be used for querying the status of the import request though invoking the following method.
- **/linkeddata/ld/ldupdate**: This method is used for performing incremental updates to the LD data already imported through SPARUL update statements. In this way, the data provider will not only be able to store his/her data sets in a LD form but also update them when needed (when such data change or new data are generated). Obviously, this method should be used in conjunction with the first one according to the direct import mode where the data provider has direct control of the LD created or updated corresponding to his/her datasets.
- **/linkeddata/ld/import_status**: This method takes the import call ID and returns the status of the LD importing for non-blocking import calls.
- **/linkeddata/ld/addR2RMLMappings**: This method realizes the (basic) indirect import mode by enabling the data provider to provide an R2RML specification which specifies the mapping between his/her relational data to LD. As soon as the respective function of the underlying triple-store is called, which successfully starts the mapping procedure, the method returns a successful message.
- **/linkeddata/ld/addXSLMappings**: this method realizes the alternative indirect import mode by enabling the data provider to provide a XSLT mapping specification on how his/her XML-based data will be transformed to LD. In contrast to the previous method, an answer is returned to the user only when the mapping has been performed and the respective LD has been successfully stored in the underlying triple-store.

Function	Description	Documentation
FTP Data Upload	Allows to upload data in the INGC platform through an FTP connection.	See text above
HTTP Data Upload	Allowsto upload data in the INGC platform through an HTTP connection.	See text above
DB scripts Upload and execution	Allows passing SQL-scripts for management of databases.	See text above
Linked Data (Direct) Upload	Allows uploading and storing LD (corresponding to the provider's data sets) into the underlying triplestore.	See text above + the API documentation.
Linked Data Update	Allows updating the LD stored through any type of modification (e.g., update of existing LD or addition of new LD)	See text above + the API documentation
Relational to Linked Data Mapping	Allows specifying the way the data provider's relational data can be mapped to the LD to be generated and stored in the underlying triple-store.	Once the mapping is established, not only the LD are generated and stored but also updated based on the respective modifications of the relational data. See also API documentation. Please note that currently the relational data should be stored in the Virtuoso's native database so as to enable this mapping based on the actual version exploited for this triple-store.
XML to Linked Data Mapping	Allows specifying the way the data provider's XML data can be mapped to the LD to be generated and stored in the underlying triple-store.	The provider should include in a zip file inline in the request his/her XML data sets. See also API documentation.

2.1.3.3. Concluding Remarks On InGeoCloudS Integration Work

The *DataImport* service implementation in Pilot2 currently answers use cases- related applications and data provider's workflows for ensuring synchronisation of data produced in premises and production of final datasets in the Cloud. Nevertheless, we can envisage extending the import and synchronisation facilities in order to answer broader needs and constraints that new data providers could express:

- HTTPS imports with possible drag&drop facilities in the Web interface
- Harvesting of catalogs

- Diagnostics tools concerning validity of R2RML mapping files

On the other hand, the *Linked Data Management* service can be used to either directly import LD that have been generated by the data providers or indirectly by specifying mappings from relational or XML data to LD. Thus, it can be considered as an alternative implementation of an import service, catering for all possible use cases, which enables integrating the provider's data sets with the data sets already stored in the same or different thematic field (based on the GSOM model) in such a way that the potential applications/use cases can benefit from all of these (linked) data through performing the respective cross-provider or cross-thematic SPARQL queries (apart from the provider-specific ones). The Linked Data Management service is currently exploited by the use cases in Section 2.2.5 (Ground Water Resource Management in Granular Aquifers) and in Section 2.2.6 (Landslide Use Case).

2.1.4. DATA PUBLICATION SERVICES

These services as described in [R3] and gathers two major functions in relation with the management of spatial datasets:

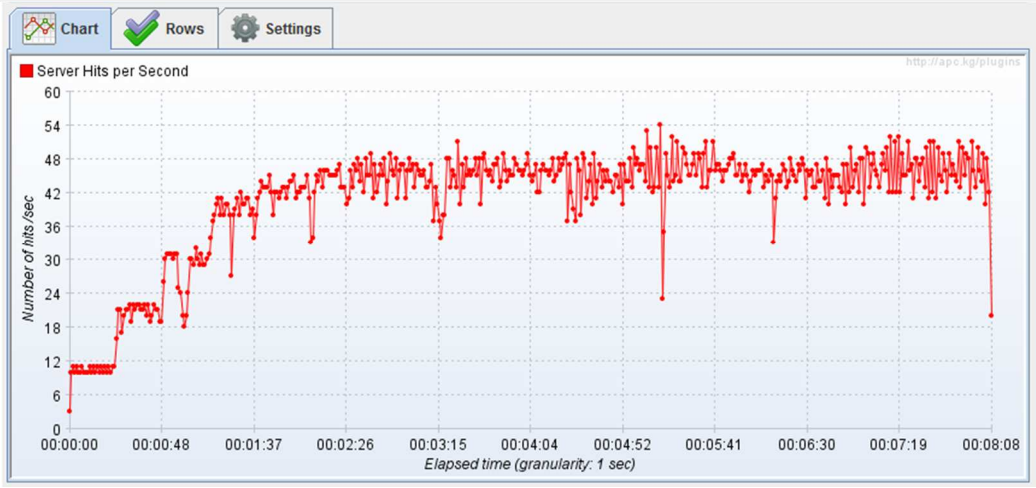
- creation and access of metadata provided by a catalog of (geo) metadata with the use of CSW standards ;
- creation and access of geospatial services defined by OGC: WMS and WFS.

The access services should be in compliance with the INSPIRE guidelines for the discovery, view and download network services.

2.1.4.1. Mapping From Use Case Requirements to Technical Requirements

#rq	Description	Evaluation
Table 2#5	Allow to comply with INSPIRE performance requirements for services and guarantee the dissemination of my user without worrying about performance and scalability issues	INSPIRE technical performance (99 % available, performance and Capacity)[3]
	Resulting technical requirements and technical choices	
	The pilot 2 is defined and deployed in a very scalable solution with the use of the Amazon cloud infrastructure. The INSPIRE service, in particular for view and download services, are deployed in the data publication component (mapserver) in an auto-scalable group of servers. The first testbeds have been implemented for the WMS service. With a large instance, the number of simultaneous requests of GetMap operation could peak at 50 requests (INSPIRE requirements of 20 requests in less 5 seconds).	

[3] For more details, our Wiki collects all considerations under http://www.ingeoclouds.eu/?q=inspire-performance-requirements

#rq	Description	Evaluation
	 *Figure 1: Technical Test with a WMS GetMap request (800 Ko of a image) in a large instance for data publication component* The supervision of the infrastructure should allow managing the quality of each service and provides an efficient service with the objective of 99 % of availability.	

#rq	Description	Evaluation
Table 3#3	Facility to re-use the Susceptibility map without wasting time to distribute to support team (disk storage distribution,…).	Web services to re-use Susceptibility map as WMS and download
	Resulting technical requirements and technical choices	
	The data publication service allows to create Web services for susceptibility requirements and to be re-usable by users. The implementation is in progress to manage correctly specific formats required by the landslides use case.	

#rq	Description	Evaluation
Table 4#5, table 6#5 and table 7#8	Facilitate the re-use of shakemap data Facility of re-use for: contour maps, water table contour maps, water table and piezometric maps, hydraulic conductivity and transmissivity values plots, iso-chemical maps of various chemical compounds, salinization maps and groundwater pollution/contamination maps without use of support team	WMS and WFS services will be used to give global remote access to the data
	Resulting technical requirements and technical choices	
	The data publication service allows to create Web services for shakemaps, groundwater resources and active landslides requirements and to be re-usable by users and applications.	

2.1.4.2. Description of Functionalities implementation

The data publication service is in charge of the creation, publication and access geo- datasets with the following functionalities:

Function	Description	Documentation
Visualize geodata	Allows to visualize a dataset as an image or a set of datasets	The service provides a WMS service for all datasets included in the INGC middleware (elastic file server or elastic database server)
Access geodata	Allows to interact with datasets	The service provides a WFS service for all datasets included in the INGC middleware
Download geodata	Allows to create HTTP downloads of files	The service provides an ATOM service (as required by INSPIRE).

The data publication services is also in charge of the management of metadata compliance with international standards for metadata (ISO 19*) and INSPIRE.

Function	Description	Documentation
Create metadata	Allows to create metadata for datasets, services, maps,…	The services provides a CSW-T (transactional) to create new metadata
Search and access metadata	Allows to search and access metadata included in the INGC catalog	The service provides a “classic” CSW to search and access metadata in Dublin Core or ISO19139 standards.

The Linked Data Management Service also provides particular functions which enable exporting either LD specifications of existing data sets or geo-data files which conform to particular standard or standardized formats, such as KML, Shape and GeoJSON. It also provides a function which allows exporting XML-based INSPIRE-compliant data from the existing data sets stored in a LD form. These functions are described below:

Function	Description	Documentation
Export LD	Allows exporting LD about (scientific) data sets in particular RDF formats.	The requester should indicate whether the data for a specific provider should be exported and/or data which conform to particular triple patterns (e.g., mapping to the description of particular scientific resources – this enables to gather data for many data providers and not just one). See also the API documentation (**/linkeddata/ld/ldexport**).
Export Geo-Data files	Allows the user to transform the results of an arbitrary geo-related SPARQL query to a geo-data file which conforms to particular standard or standardized geo-	The restriction of using this service is that the SPARQL query results should contain geo-spatial information (i.e., WML or WKT geometry specifications). The current output file representation formats supported are: GeoJSON, GML, KML, and Shape. See also the API documentation (**/linkeddata/geold/geoldtransform**).

	formats.	
Export Inspire-Compliant data	Allows exporting XML-based INSPIRE-compliant data from the data sets stored in the infrastructure in LD form.	The user can select from which thematic field to download the related data as well as the data location (e.g., Greece or Denmark). The user has also the option to download all data for any country. Please note that not all information pertaining to the data sets stored is provided due to current mapping from GSOM to INSPIRE as well as INSPIRE incapability to completely represent particular types of information. See also the API documentation (**/linkeddata/ld/inspire_export**)

2.1.4.3. Integration within the INGC Infrastructure

The data publication service has been deployed as a service acting between the middleware part of the architecture (Elastic file server, database server…) and the portal and tools. As described in the D3.2 – 3.1.4 and 3.1.7, the service is provided by the elastic geospatial server for WMS/WFS and by Geonetwork application for the metadata catalog /CSW.

The integration in the INGC infrastructure was done from the pilot 1 and has been improved in the pilot 2, in particular for the management of the scalability of the elastic geospatial server (auto scaling).

All services provided by this components are available for each data provider and application provider to use it in the geo-application (portal…) and others re-uses.

The respective functions of the Linked Data Management service are provided for all users and can be used for exporting: (a) LD about any scientific data resource, (b) geo-spatial files for the whole or parts of the data sets stored/imported by the data providers (this is the capability of transforming arbitrary SPARQL queries to feature collection representations), and (c) XML-based INSPIRE-compliant data for all data sets or data sets corresponding to specific thematic fields and locations. Thus, they can be used by any use case or application which requires such functionality. The LD export function is available from pilot 1 but has been modified in pilot 2 so as to return inline in the response the exported LD file, while the rest export functions are available from pilot 2.

2.1.4.4. Concluding Remarks on InGeoCloudS Integration Work

The data publication service implementation in Pilot2 currently answers use cases- create geoservices and access geodata – and was deployed without major difficulties in the cloud infrastructure. The progress work is currently the improvement of the performance of the view service with the deployment of a cache (tiled cache) dedicated to some large datasets. The cache system requires new technical components, in particular new storage management for a great number of files (millions of files) and synchronization process to create cache systems.

The respective functions of the Linked Data Management service can also be used for providing access to geo-data for any possible use case / application. They have been easily developed and deployed in the INGC infrastructure and have used open-source APIs and tools, such as geo-tools and the Sesame API. Especially, the recently –completed export function enables the compliance to INSPIRE directives, thus

providing added-value functionality for the InGeoCloudS project as well as the potential new users of the project's platform. Currently, it is investigated whether specific extensions can be performed to these functions, such as respecting the licences for the respective data sets to be exported, supporting other geo-spatial feature collection representation formats, and expanding the INSPIRE-compliant export functionality for the results of arbitrary SPARQL queries.

2.2. USE CASE SPECIFIC SERVICES

2.2.1. DATA PUBLICATION USE CASE (UC1 – WAS DOW/UC5)

This use case gathers the publication of geospatial datasets as required by INSPIRE and the creation of maps on the Internet. As described in the DOW, some technical components were developed by the BRGM and national ministry of Environment (project Carmen, project GeoSource) for national usage and hosted in local web infrastructure. The objectives of this use case were to create a "European" service for the publication and map creation in a scalable (cloud) infrastructure.

2.2.1.1. Mapping From Use Case requirements To Technical Requirements

#rq	Description	Evaluation
1	Publish maps on the web to disseminate a "message" in relation with the organisation policy: shakemap access, geology maps,…	Web map available in the INGC for my information
	Resulting technical requirements and technical choices	
	The publication service allows creating any kind of maps and publishing datasets as services. Maps are created with the backOffice interface for data provider and are available for public with the frontOffice solution. All tools and software are deployed in the cloud infrastructure on the top of the INGC infrastructure (data publication, data import, elastic FS, elastic DB,…).	
#rq	Description	Evaluation
2	Facilitate the re-use of my data without wasting of time distributing my datasets to a support team (disk storage distribution,…). The cloud service should allow to transform the "accessibility information" requirements into a "direct" dissemination strategy with web services	Web services to re-use my dataset as WMS, WFS and download
	Resulting technical requirements and technical choices	
	All maps and datasets could be published as web services defined by the data provider : WMS, WFS and Download services are available with the use of the data publication component (mapserver) and documentation of this services are published in the catalog.	
#rq	Description	Evaluation
3	Have a catalogue of my data and my services compliant with INSPIRE requirements (regulation) so as to be referenced in the scientific community (like INSPIRE, GEOSS, national catalogue,…)	Catalogue of my data and my services available
	Resulting technical requirements and technical choices	
	A unique catalogue for all data provider was published in the pilot 2 of the INGC to describe datasets, series of datasets, services and maps. The catalogue is deployed with the use of Geonetwork solution hosted in the INGC infrastructure with PostGres database (elastic DB). A specific development was added in the backOffice interface to facilitate the link between the map creation and the metadata creation and edition.	

#rq	Description	Evaluation
4	Have visualisation and download services as required by INSPIRE and propose harmonized datasets defined by INSPIRE	INSPIRE compliant web services and INSPIRE data model
	Resulting technical requirements and technical choices	
	The pilot 2 allows to create a first version of INSPIRE web services in the data publication components (mapserver and geonetwork). The WMS, WFS and CSW are available for the re-use of datasets. The WMS, CSW and ATOM service are compliance with the INSPIRE guidance, the WFS are partially compliance (in waiting for the implementation of the version 2 of the WFS service in the opensource MapServer). The INSPIRE data model compliance is not currently implemented. The strategy will be to use the linked data module (GSOM) to interact with INSPIRE data model.	
6	Allow access to valuable web services matching my needs: • Integration of services in a web portal like OneGeology portal, ewater portal, or INSPIRE geoportal,… • Integration of data or services published by the INGC in my application	Re-use of provider's data by end-users or other projects in charge of portals,…
	Resulting technical requirements and technical choices	
	The architecture defined for the publication use case provides a set of functionalities to create maps on Internet and to reuse the maps and the datasets in others projects : - direct URL to public maps are available and could be integrate easily in a other project or websites ; - URL for view and download services are also provided in the catalog and could be integrated in a other map or other webmapping project.	

2.2.1.2. Description of Functionalities / Implementation

As described in the D4.1, the use case is composed of two domains: a) publication / map edition for data provider (named Backoffice) and b) visualization of results (datasets and maps) and reuse with INSPIRE services (named FrontOffice). The description of functionalities is in the following tables:

2.2.1.2.1. GIS Client functions for data visualization in the FrontOffice (and OGC services)

Function	Description	Documentation
Zoom and display all active layers	Draw rectangle on map when user drags with left-mouse button pressed, Redraw map zoomed to selected area.	Functionalities available in the frontoffice and with the specific button "Zoom + ". The user can choose which layers must be displayed in the "layers" module. The WMS service (GetMap operation) allows to interact with a map/dataset to select the

		area.
Zoom-out and display all active layers	Draw rectangle on map when user drags with left-mouse button pressed, Redraw map zoomed-out to area reflecting size of area the user drew (small area drawn = high zoom-out factor)	Functionalities available in the frontoffice and with the specific button “Zoom - ”. The user can choose which layers must be displayed in the “layers” module. The WMS service (GetMap operation) allows to interact with a map/dataset to select the area.
Pan and display all active layers	Move origin of map when user drags with left-mouse button pressed, Redraw map with new centre.	Functionalities available in the frontoffice and with the specific button “Pan”. The WMS service (GetMap operation) allows interacting with a map/dataset to select the area.
Select rectangular area	Draw rectangle on map when user drags with left-mouse button pressed	Functionalities available in the frontoffice and with the specific button “Select”. The functionality is extended to select with a circle and a general polygon. The WFS service (GetFeature) allows to request datasets in a specific area (bounding Box).
Display background map	Redraw map which has been selected by user	Included in the frontOffice and the backOffice Available as a WMS service in the INGC infrastructure.
Select map projection	Redraw map projected according to specified map projection	Not available in the frontOffice interfacebut the user could define the type of coordinates in the tracker (position of the mouse in X/Y – Long/Lat). The WMS and WFS service allows to define map projection with the use of EPSG code.
Highlight rectangular area	Draw rectangle in area	The select functionality allows highlighting a area in the frontOffice. The WFS service allows to select a area.
Add list to dropdown box	Add entries to a dropdown box on user interface	Not available
Display the legend	Show the legend of each layer	If defined by the dataprovider, the module “Legend” is available in the interface. The WMS service (GetGraphicLegend) allows displaying the legend of the service.
Display the overview	Show the overview of	If defined by the data provider in the

map	the map in a border of the map	backOffice, the module "OverView" is available in the frontOffice interface.
Display coordinates	Show the X/Y (or long/lat) of the tracker	The tracker is available in the frontOffice interface.
Go To…	Allow to zoom in function of a textual criteria	The GoTo functionality is available in the frontOffice after configuration by the data provider.
Print the map	Allow to print and export as image a map	Print the map is available in the frontOffice interface.
Request objects in the layers	Allow to request all active layers after selecting a area on the map. Provide all attribute information in the data (DB, FS)	A request module is available in the frontOffice interface allowing to search / find / zoom on specific objects.
Fetch list	Fetch a list of values for e.g. a code list from Elastic DBMS	Not available
Fetch background map	Fetch a map layer from Elastic File Server or Elastic DBMS	Background maps are made available in the infrastructure.
Stream data	Stream a collection of objects to the user	Not available.

2.2.1.2.2. Specific functions associated to the map access

Function	Description	Documentation
Add annotations	Allow to a user to create annotations of the map : points, lines, polygons, text	The annotation panel is available in the frontoffice.
Search in the data	Allow to define a specific search in a layer and to directly zoom on the map on the results (geometry objects)	Available.

2.2.1.2.3. Map Editor (backoffice)

Function	Description	Documentation
Authenticate in the map web editor	Function to provide only a access for a authenticated author	Pilot 1 and 2 requires to login with a specific interface. A updated version allows to use the SSO service (authentication at the portal level only)
Create a map	Function to create a map from a template or a "null" map.	The backoffice allows creating a map with his own datasets and customizing the map.
Add and organize layers	Function gathering all	The backoffice provides all functionalities to

to a map	operations to add a layer to the map, gives a name/alias to the layer + organizes layers (order). Layers datasources are data files (File System) or DB PostGres Table/Views or background layer provided by INGC or external WMS/WFS.	organize the datasets and layers in the map.
Add symbology to a layer	Allow to change the symbology of a layer in the map color, text symbols, distribution of data,…	The backOffice provides different modules to customize each dataset and create specific symbology.
Organize the legend	Allow to organize the legend in the map : order of the layer, collapsing layers in group	The backOffice allows organizing the legend in the map.
Add functions to the map	Allow to active some functions available in the public viewer : type of buttons, scale,	The backOffice allows to configure each functionalities available in the front Office: Download options, print, buttons,…
Choose the logo of the map	Allow to add/modify the logo and header of the map in the public viewer	The header of the map could be customize with specific logo.
Publish the map	Record the map and publish automatically the map for a public use	The backOffice allows to create a public map with a specific address.

2.2.1.2.4. Datasets Download

Function	Description	Documentation
Download data from Maps	Function to download datasets available in the map if the author authorizes it: a function in the portal with a button “Download” and some form to fill.	The download service is available in the interface for the pilot 2 and provides all options to download datasets published by the provider. The WFS service and the ATOM service (“download” services) are also included in the pilot 2 of the INGC.

2.2.1.2.5. INSPIRE Processes

Function	Description	Documentation
Provide data as service	The author has the capability to decide if data has a public OGC service address: WMS	The backOffice allows to create a view service and a download service in regard of the INSPIRE guidance.

	and WFS. The service published should be compliance with INSPIRE requirements	The view service is 100 % compliance with INSPIRE requirements. The download service is compliance with ATOM requirements defined in the INSPIRE guidance. The WFS service is not fully compliance with INSPIRE guidance.
Check data in regard of data models	Service to validate user's data are compliant or not in regard of a "data model" template (for example INSPIRE model or linked-data model,…).	Not implemented
Metadata management	Functions to publish metadata for each layer (=dataset) with compliance with INSPIRE requirements	A metadata catalogue is available in the pilot 2 of the INGC with a Geonetwork solution. The catalogue includes the description of datasets, series of datasets, services and map. The catalogue is compliance with INSPIRE guidance. A CSW service is available for the reuse and compliance with the discovery service defined by INSPIRE.

2.2.1.3. Integration within the INGC Infrastructure

The use case is fully integrated in the INGC infrastructure to avoid to re-develop some common components (import, etc.) and to use the scalability of the INGC infrastructure to fulfil some technical requirements. The following figure describes main components used for the use case and the motivation of using this technical infrastructure.

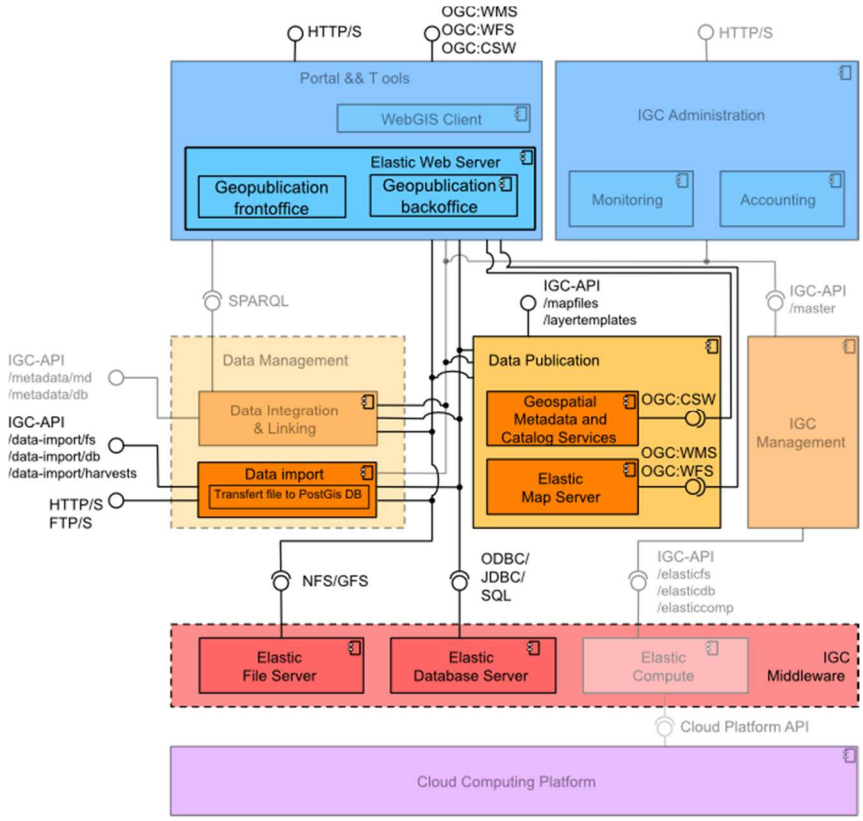


Figure 2: INGC infrastructure used by the GeoPublication UC

Figure 2 shows where the different services supporting the GeoPublication Use Case are located in the INGC system and which interfaces they use. The services are grouped into the following:

- Elastic File Server: Allows to store all datasets, PDF, configuration files for each data provider (account) used by the back and the front Office of the GeoPublication application. Direct link with NFS solution.
- Elastic DB: Allows storing (optionaly) geodatasets in the DB proposed in the infrastructure (PostGres + PostGIS module). Link with JDBC.
- Data Management / Data Import: Provide solution to push datasets in the INGC infrastructure with the use of FTP protocol between local organisation and INGC. The UC added a module to transfer/import flat files in the DB with synchronization mechanism.
- Data Publication: Use the data publication API to create metadata, create WMS and WFS services and download services. The data publication is used by the front (readonly) and the back Office (CRUD). Use the API for the edition of the configuration of services and OGC services for the access.
- Portal and Tools: The application GeoPublication is deployed in the portal component in a specific server.

2.2.1.4. Influence on the design of the INGC resulting from the Use Case Implementation

The GeoPublication use case's implementation influenced the InGeoCloudS design choices in the following ways:

- The functional objectives of the use case were split in two «components»: generic (and reusable) functionalities have been integrated in the data publication component and specific functionalities

and Web Interfaces have been deployed in the specific portal for the UC. The use case allowed to define requirements for the data publication and to specify web services and IT interfaces.

- The elastic database has been configured properly to manage geospatial datasets and tables with the additional extension PostGis on Postgres DB.

2.2.1.5. Concluding Remarks On InGeoCloudS Integration Work

The integration of the GeoPublication use case in the INGC infrastructure was done without major difficulties but some challenges were identified during the migration process in the cloud:

- A split in components of the current application was done to be compliance with the INGC infrastructure, in particular to optimize the scalability of the application;
- The code (in php) of the current application was modified without difficulties to use the API defined by INGC but some versions of softwares (MapServer, PostGIS,...), not compatible with the current application, required lot of changes and regression tests before the deployment ;
- the exploitation of the UC is in its early stages but it is showing promise for the scalability of the cloud and the performance that the architecture will garantee. The capability to create more and more maps, publish more and more datasets and services, add new accounts without impact for the others data providers / users is very useful – except the budget required to run all the configuration.

2.2.2. SUSCEPTIBILITY MAPS USE CASE (UC2 – WAS DOW/UC2)

2.2.2.1. Mapping From Use Case Requirements to Technical Requirements

#rq	Description	Evaluation
1	Pushing of raw data and the execution of the calculation of Susceptibility map of triggering landslides due to rainfall forecast must fulfil the following requirements: 1. Be automatic 2. Be executed every day with new precipitation forecast	The process is automatic and executed every day
	Resulting technical requirements and technical choices	
	GeoZS have created a service which runs daily and updates the database with a new data. The first part is a locally hosted service, which collects precipitation forecast data from Slovenian Environment Agency's ftp server. The second part is a calculation service running at INGC, which fetches the records from the first service and performs the "Susceptibility map of triggering landslides due to rainfall forecast" calculations and necessary database operations (INSERT into database). The second part was uploaded to the INGC platform using the INGC API Data Import Service and is scheduled using the INGC Scheduling service..	
#rq	Description	Evaluation
2	Dissemination of Susceptibility map of triggering landslides due to rainfall forecast	Web map available in the INGC for my information
	Resulting technical requirements and technical choices	

#rq	Description	Evaluation
	In order for the web GIS (SiTools) to access the GeoZS data, we had to implement a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database….	
#rq	Description	Evaluation
3	Facility to re-use the Susceptibility map without wasting time to distribute to support team (disk storage distribution,…).	Web services to re-use Susceptibility map as WMS and download.
	Resulting technical requirements and technical choices	
	GeoZS implemented a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database every day	
#rq	Description	Evaluation
4	Susceptibility map of triggering landslides due to rainfall forecast in INGC system must be available and must be clearly marked with the date and time of the calculation and validity.	The map is available every day.
	Resulting technical requirements and technical choices	
	WMS and WFS service are exposed by map server	
#rq	Description	Evaluation
5	Web-application must contain a map with the following functionalities: 1. Zoom in and out. 2. Pan. 3. Display topographical background map. 4. Display other background maps at user's choice including Corine Land Cover and OneGeologyEurope surface geology. 5. Enable user to select and display other maps available as WMS/WFS. 6. Display of Susceptibility map 7. Display of the legend for the Susceptibility map. 8. Show information about Susceptibility map when location is clicked. 9. Possibility of setting transparency All maps must be redrawn within 5 seconds after a change in map area.	.Functions are available
	Resulting technical requirements and technical choices	

#rq	Description	Evaluation
	WebGIS Client accessing the GeoZS data with WMS/WFS	
#rq	Description	Evaluation
6	Integration of services in a web portal of Administration of the Republic of Slovenia for civil protection and disaster relief to inform them of an increased landslide hazard as a consequence of heavy precipitation and enable them to mitigate risks.	First responders will have an early warning system that will give them enough time to save lives, property and also predict negative effects of landslides on infrastructure.
	Resulting technical requirements and technical choices	
	GeoZS implement a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database in INGC. These services will be implemented in the Administration of the Republic of Slovenia for civil protection and disaster relief system.	
#rq	Description	Evaluation
7	Integration of services in a web portal of Environmental Agency of the Republic of Slovenia (ARSO).	ARSO primary mission is to monitor, analyse and forecast natural phenomena and processes in the environment, and to reduce natural threats to people and property
	Resulting technical requirements and technical choices	
	GeoZS implement a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database. These services will be implemented in the Environmental Agency of the Republic of Slovenia portal.	
#rq	Description	Evaluation
8-13	Susceptibility map of triggering landslides due to rainfall forecast is available for general public and different users (planners, insures and risk managers, real estate agent, Infrastructure owners and operators, municipalities/local authorities, …).	The map is available every day
	Resulting technical requirements and technical choices	
	In order for the WebGIS Client to access the GeoZS data, we had to implement a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database in INGC	

2.2.2.2. Description of Functionalities implementation

Function	Description	Documentation
Precipitation forecast data upload	Collects precipitation forecast data from Slovenian Environment Agency's ftp server	In order to support this functionality, we implemented a service which runs daily and updates the database with a new precipitation forecast data. This locally hosted service collects precipitation forecast data from Slovenian Environment Agency ftp server. The data upload triggers the calculation and publication of a new Susceptibility map

Susceptibility map Computation	Precipitation forecast data uploading triggers the process of calculating a new Susceptibility map. Process runs automatically every day.	In order to support this functionality, we implemented a service which fetches the records from the "Precipitation forecast data upload" service and two static maps and performs the "Susceptibility map of triggering landslides due to rainfall forecast" calculations and necessary database operations (INSERT into database)The process of Susceptibility map calculation is triggered by the "Data Upload" process.
Display Susceptibility map for specific area	Displays a Susceptibility map in selected area.	The Web GIS client is using SiTools for displaying the maps.

2.2.2.3. Integration with the InGeoCloudS Infrastructure

a) The Susceptibility map use case was not implemented within the GeoZS infrastructure before the InGeoCloudS project. It was implemented during the InGeoCloudS project.
b) The implementation of the GeoZS use case within the InGeoCloudS infrastructure impacted on the following components: ElasticDB, ElasticFS, Apache Web Server, Map server etc. Creation of the database, inserting values into database from different sources, services for automatic calculation and prediction.
c) In order to support Susceptibility map use case within INGC, we had to import the GeoZS Susceptibility map database into the INGC platform. The initial import was done in four steps:
    1. Create a local PostgreSQL database with the GeoZS data-model and a snapshot of the database.
    2. Create GeoZS as a provider within INGC and getting the DBMS connection info through the INGC API Master Service and Elastic Database Service.
    3. Create the same tables of the local PostgreSQL database into the INGC platform.
    4. Insert the last version of the 2 static tables (Landslide triggering threshold values and landslide susceptibility map) in the INGC GeoZSpostgreSQL database. Other tables will be automatically inserted with the daily "Susceptibility map of triggering landslides due to rainfall forecast" calculation service.

In order to keep the data within the PostgreSQL database up-to-date, we have created a service which runs daily and updates the database with a new data. The first part is a locally hosted service, which collects precipitation forecast data from Slovenian Environment Agency ftp server. The second part is a provider "Susceptibility map of triggering landslides due to rainfall forecast" calculation service running at INGC, which fetches the records from the first service and performs the "Susceptibility map of triggering landslides due to rainfall forecast" calculations and necessary database operations (INSERT into database). The second part was uploaded to the INGC platform using the INGC API Data Import Service and is scheduled using the INGC Scheduling service.
In order for the web GISClient to access the GeoZS data, we had to implement a service (with map server) which creates WMS service from our PostgreSQL / PostGIS database.

d) Web GIS client for displaying the Susceptibility map of triggering landslides due to rainfall forecast data.

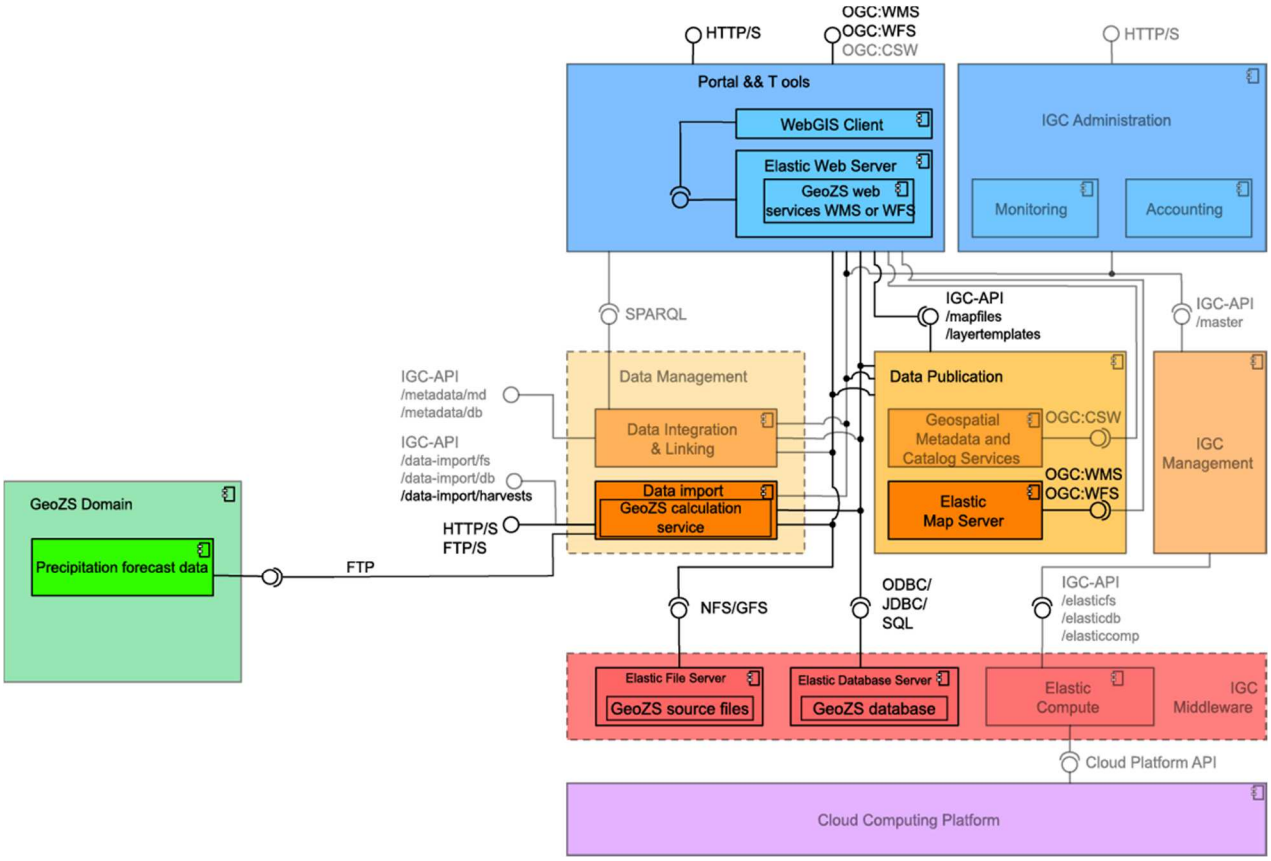


Figure 3: Services for Susceptibilities Use Case in the INGC system

Figure 3 shows where the different services supporting the Susceptibilities Use Case are located in the INGC system and which interfaces they use. The services are grouped into the following:

- GeoZS web services WMS of WFS: These services are exposed to web GIS client to display GEOZS susceptibility map and to others who wants to download them as WMS or WFS to use them in other applications
- GeoZS calculation service: This service uses the data from the GeoZS database and Precipitation forecast data and calculates susceptibility map of triggering landslides due to rainfall forecast
- GeoZS source files: These are all the files that GeoZS use case needs to work: executable jar files, log files and the script that executes the jar files.
- GeoZS database: Here resides the PostgreSQL database containing the GeoZS data on the INGC system.
- Precipitation forecast data: These are weather forecast data that we fetch from FTP server and transform them into GRID format and store them into GeoZS database for GeoZS calculation service.

2.2.2.4. Influence on the design of the INGC resulting from the Use Case Implementation

None

2.2.2.5. Concluding Remarks On InGeoCloudS Integration Work

Overall, the InGeoCloudS platform has provided or will provide according to planned schedule, all services that are needed by the GeoZS use case in order to be fully functional. The few problems encountered during integration of GeoZS use case only came from the parallel development of the INGC platform.

2.2.3. SHAKE-MAPS USE CASE (UC3 – WAS DOW/UC1)

This use case was not implemented on the data providers (EPPO's) infrastructure prior to the InGeoCloudS project, so all services are newly developed.

2.2.3.1. Mapping from Use Case requirements to Technical Requirements

#rq	Description	Evaluation
1	Automatic push of data and calculation of a shakemap in almost real-time for new earthquake events	The process will be automatic and executed each time a new event occurs
	Resulting technical requirements and technical choices	
	A fully automated process is triggered at EPPO local infrastructure each time a new earthquake event occurs. This process utilizes the Data Import and Geo-Computation APIs and a) pushes data in the InGeoCloudS platform, b) initiates a geo-processing instance which calculates a new shakemap. The whole process is finished within a few minutes, thus providing results in almost real-time as required by the application.	
#rq	Description	Evaluation
2	On demand push of data and calculation of a shakemap for earthquake scenarios or old earthquakes	A WPS Service will be provided to push data and ask for a shakemap calculation
	Resulting technical requirements and technical choices	
	This requirement is still under implementation.	
#rq	Description	Evaluation
3	Make shakemap data publicly available for download in various formats	Data will be made available for download through a web published chronological archive of earthquakes
	Resulting technical requirements and technical choices	
	As soon as the new shakemap is calculated all data are stored in the Elastic File Server in EPPO's private workspace and are automatically published in a web chronological archive (html based) through the Elastic Web Server. Users can then browse / view and download all or part of the desired data.	
#rq	Description	Evaluation

#rq	Description	Evaluation
4	Publish shakemap as a geo-referenced map over an earth background layer	A GIS Service will be used to publish map
	Resulting technical requirements and technical choices	
	This requirement is still under implementation.	
#rq	Description	Evaluation
5	Facilitate the re-use of shakemap data	WMS and WFS services will be used to give global remote access to the data
	Resulting technical requirements and technical choices	
	This requirement is still under implementation.	
#rq	Description	Evaluation
6	Send notifications to interested actors according to their personalized preferences	The system will send notifications to registered users
	Resulting technical requirements and technical choices	
	This service is part of the Administrative Services (see also 2.1.1 Administrative (User management) services) and is not yet implemented: some generalisation of the function is under discussion and shall be introduced to the platform yet in 2013.	
#rq	Description	Evaluation
7	View and download shakemap data	Shakemap data will be available for download after an earthquake event
	Resulting technical requirements and technical choices	
	All shakemap data are published in the web by the InGeoCloudS Elastic Web Server, through an easy to use web interface. Functionalities such as browse, view and download shakemaps are available.	
#rq	Description	Evaluation
8	Search / filter shakemap data	A search / filtering functionality will be available
	Resulting technical requirements and technical choices	
	This requirement is still under implementation.	
#rq	Description	Evaluation

#rq	Description	Evaluation
9	Request notification on new shakemaps	Registered users will be able to subscribe to event notifications according to specific criteria
	Resulting technical requirements and technical choices	
	Same as requirement #6.	

2.2.3.2. Description of Functionalities Implementation

Function	Description	Documentation
Shake-Map Computation	Data uploading triggers the process of calculating a new shake-map. Process runs with data provider's credentials.	To support the automatic calculation of new shakemaps we implemented a geo-processing service. Each time new data are available this service is called using the Elastic Compute API methods to start/stop a new computation instance.
FTP-enabled Shake-Map download	A user accesses shake-map archive by FTP and requests to download data: in practice, the user is provided with an FTP server, and asks the system to move the data into the owned ftp server.	Authenticated users can be provided through the DataImport API with FTP server access facilities in order to access their private data.
Shake-Map Data	A user browses shake-map archive over HTTP	A web interface is published through the Elastic Web Server. A chronological archive of shakemaps is available for browsing through all shakemap data.
HTTP-enabled Shake-Map download	A user accesses shake-map archive by HTTP and requests to download data	The user can browse the shakemaps archive to find the desired shakemap and then download all or some of the related data.
Search Shake-Map Data	A form is provided to the user to search on shake-map metadata	Not implemented yet
View shake-map data for an area on map	A user requests to see a map with shake-map positions	The WebGIS interface functionality is not used yet
Fetch shake-map data	Fetch a collection of shake-map objects adhering to filter criteria	This service is related to the web GIS interface (previous functionality) and so it is not implemented yet.
Render shake-map data	Draw positions of fetched shake-map objects on map	Same as above
Display list of shake-	Display a list of shake-	This functionality will be available from both the

maps	maps incl. a link to a page showing shake-map details	simple web interface (HTML page) and from the WebGIS interface. For the former see function "Shake-map Data". The latter is not implemented yet.
Display shake-map details	Display all relevant information for a shake-map (see also "View shake-map web page" function)	This service is related to the web GIS interface and so it is not implemented yet.
View shake-map layer	User selects a shake-map from the list and requests to display it as a layer over background map	Same as above
Provide shake-map calculation as a service	The data provider can use an OGC WPS service to upload data and run the shake-map calculation process	Not implemented yet

2.2.3.3. Integration within the INGC Infrastructure

In order to support the Shakemaps Use Case integration within InGeoCloudS we needed the following:

1. A mechanism to automatically upload data each time a new earthquake event occurs. This mechanism was provided by the Data Import Service. From EPPO local facilities we call Data Import API functions that allow us to upload data in the InGeoCloudS platform.

2. A private file space to store this data accessible both from the outside world and from within InGeoCloudS. As a data / application provider, EPPO has access to a separate file space which is accessible both through the Data Import Service and directly from within INGC with common file system access protocols. When data is uploaded it can be access by the rest of INGC Services including the Geo-computational Service.

3. A mechanism to automatically start the shakemap computation when new data are available. This mechanism was provided by the Geo-computational Service. As soon as the data is uploaded in INGC, a call to Geo-computational API functions initiates a geo-processing instance which in turn performs the shakemap calculation. As soon as the calculation is over, the instance is terminated freeing thus computational resources.

4. A private file space to store the results of the computation, accessible from outside through HTTP/FTP protocols. The results of the shakemap calculation are stored again in EPPO's private file space. A specific configuration has been applied to the Elastic Web Service so that new data are automatically published to the web without manual intervention.

5. State preservation for the shakemap calculations. The shakemaps process requires some state preservation, in order to run properly. We overcame this by using again EPPO's private file space to store the status of the computations.

6. Integration with the RDF data has already been realized through defining a XSLT mapping from the original data model to GSOM and then calling the respective Linked Data Management method (/linkeddata/ld/addXSLMappings) with the original data as input. What remains to be performed is the continuous integration of new earthquake data once they are generated. To this end, it is planned to create a synchronization script which will execute the XSLT mapping specification over the new data so as to generate the respective LD and store it in the underlying Triple Store of the InGeoCloudS platform (through again the same API method). The respective Linked Data Management method (/linkeddata/ld/ldquery) can be executed in order to issue SPARQL queries and obtain back the results (see Search Shake-Map Data function) over the LD generated.

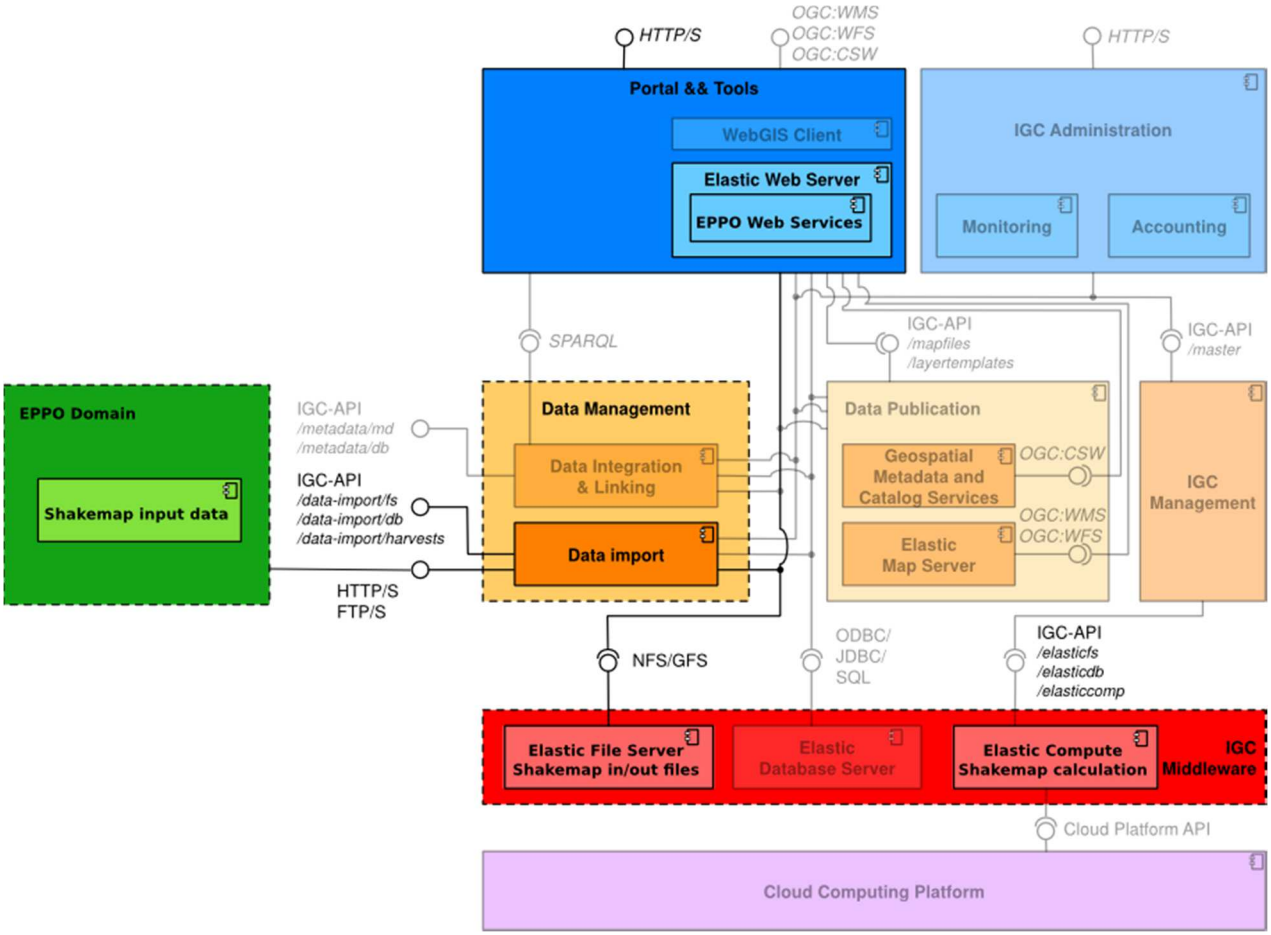


Figure 4. Services for Shakemaps Use Case in the INGC system

Figure 4 shows were the different services supporting the Shakemaps Use Case are located in the INGC system and which interfaces they use. The services are grouped into the following:

- EPPO web services: These are services directly accessible by the public. They provide browse and download functionalities for all shakemap data through direct access to the Elastic File Server.
- Data import services: We use the /data/import/fs INGC-API methods to import new earthquake data each time a new earthquake occurs. The data are imported using the FTP protocol.
- Shakemap in/out files: We use the Elastic File Server to store all shakemap related data, both the earthquake input data and the output data resulting from the shakemap calculation. Data

residing in the Elastic File Server are directly accessible from all other INGC services (Data Import, Web Services, Elastic Compute).

- Shakemap calculation: We use services exposed by the /elasticomp INGC-API to calculate a new shakemap when new earthquake data are available. The computation instance is initiated on demand and terminated when the calculation is finished.
- Earthquake data: At EPPO domain, we acquire earthquake data from a seismological network, prepare and import them in the INGC system.

2.2.3.4. Influence on the design of the INGC resulting from the Use Case Implementation

The Shakemaps use case implementation, influenced the InGeoCloudS design choices in the following ways:

1. Geo-processing API. The shakemaps was the first use case that required on demand access to computational resources.
2. The Data Import API. The shake-maps use case makes use of the API's data transfer functions to upload data directly to the Elastic File System of InGeoCloudS.
3. The Shakemap process produces html pages along with actual shakemap data. The Elastic Web Server has been configured properly to automatically publish these data.
4. The data can be accessible through an FTP Service.
5. The Notification Service was an original requirement of the Shakemaps use case.

2.2.3.5. Concluding Remarks on InGeoCloudS Integration Work

Overall, the InGeoCloudS platform has provided or will provide according to planned schedule, all services that are needed by the shakemaps use case in order to be fully functional. The integration can be considered as being relatively easy although it is not fully completed. This is due to delays in the development of the computational part of the service, which is independent of the work done in InGeoCloudS.

Since there was no previous implementation of the Shakemaps Use Case in the EPPO Domain, we didn't face any adaptation problems. As regards, the deployment/integration work a detailed list of the needs of the application as well as the solutions used to fulfill the needs is described above in 2.2.3.3 "Integration within the INGC Infrastructure".

2.2.4. PESTICIDES IN GROUNDWATER USE CASE (UC4 – WAS DOW/UC3)

This use case was not implemented on the data providers (GEUS') infrastructure prior to the InGeoCloudS project, so all services are newly developed.

2.2.4.1. Mapping From Use Case Requirements to Technical Requirements

#rq	Description	Evaluation
1	Replication from local database to INGC system must fulfil the following requirements: 1. Be automatic 2. Only include data changed since previous replication Be completed within 2 hours	Replication is available in less than 2 hours

#rq	Description	Evaluation
	Resulting technical requirements and technical choices	
	An incremental synchronization mechanism is necessary to synch the data store in the InGeoCloudS infrastructure with the data hosted at the data provider site. To this end the DataImport component was designed. It allows data providers to import data from some external repository (typically the infrastructure where this data is generated and also stored) into their dedicated workspace inside the InGeoCloudS platform. In particular, the "*harvest service*" is in charge of harvesting data located at data providers. Data providers use this service to register a harvesting mechanism that will be triggered at a defined frequency. On the data provider site the integration requires the existence of services from which the INGC can harvest the latest data.	

#rq	Description	Evaluation
2	Data in INGC system should not be more than 1 week out-of-date compared to original data in local database and must be clearly marked with last date of replication from local database.	Up-to-date data
	Resulting technical requirements and technical choices	
	The replication process must run at least once pr. week	

#rq	Description	Evaluation
3	Web-application should contain a map with the following functionalities: 1. Zoom in and out. 2. Pan. 3. Display topographical background map. 4. Display other background maps of user's choice including Corine Land Cover and OneGeologyEurope surface geology. 5. Enable user to select and display other maps available as WMS/WFS. 6. Display boreholes on the map shown with point symbols. 7. Display only boreholes where a user specified pesticide (selected from a list of values) has been found, optionally limited to water originating: a. from a certain depth interval b. from a certain lithology or lithostratigraphy (both selected from lists of values) c. from a certain time period 8. Render map in a cartographic projection suitable for the geographical area of the boreholes. 9. Show details of a borehole when this is clicked. Information must include technical completion	All functions are available.

#rq	Description	Evaluation
	of borehole, lithological and lithostratigraphic log, time series of amounts of water abstracted, concentration of pesticide, detection limits and legal limits. Map, including boreholes and background maps, must be redrawn within 5 seconds after a change in map area and/or filter criteria for boreholes.	
	Resulting technical requirements and technical choices	
	The satisfaction of these requirements needed the implementation of a general purpose Web GIS interface (see section 2.1.2). To this end, the application has been adapted to utilize a generic GeoJSON web service, which functions as a data provider for the Web GIS interface.	

#rq	Description	Evaluation
4	WMS and WFS services exposing borehole locations and results of chemical analyses.	WMS and WFS services present and documented.
	Resulting technical requirements and technical choices	
	No WMS or WFS services are exposed by the Pesticides In Groundwater use case. We use vector services (GeoJSON) instead.	

2.2.4.2. Description of Functionalities Implementation

Function	Description	Documentation
Fetch borehole data	Fetch from Elastic DBMS a collection of borehole objects adhering to filter criteria	In order to support this functionality, we implemented a web-service which takes all the filter criteria as well as the bounding box of the viewport as input parameters and returns GeoJSON data. For each borehole within the bounding box it filters using the different criteria and returns the name, unique identifier and location of the resulting boreholes.
Render borehole data	Draw positions of fetched boreholes on map	The Web GIS client is using OpenLayers for rendering the maps. OpenLayers have built in functionality to render vector layers with data coming from GeoJSON services. This draws an icon on the map using the location from the GeoJSON list. The web GIS client is appending the filter criteria to the request performed by OpenLayers to the borehole service described above.
Display list of boreholes	Display a list of boreholes incl. a link to a page showing borehole details	In order to support this functionality, we implemented I simple HTML search page in PHP, which utilizes a service similar to “Fetch borehole data” – but it returns the schema data instead of GeoJSON.
Display borehole details	Display all relevant information for a borehole including geology, time series of	In order to support this functionality, we implemented a set of services. The first service takes a borehole identifier as input parameter and returns HTML code displaying borehole information as well as links to the remaining services.

	concentrations of chemical compounds and water levels, etc.	The second service takes a borehole identifier as input parameter and returns an image of the geology through which the borehole passes. This image is generated on the fly from geology data within the groundwater database. The third service takes a borehole identifier and a list of compounds of interest as input parameters and returns an image of the time series of the measured values for the compounds within samples taken from the borehole. The fourth service takes a borehole identifier as input parameter and returns an image of the time series of the measured water levels for each intake within the borehole.
Download measurements for a borehole	Allow to download data included in the portal	In order to support this functionality, we implemented a service which takes a list of borehole identifiers as input parameter and returns a zip-file containing all relevant data from the groundwater database related to these boreholes. Each csv-file corresponds to a single table within the data model.

2.2.4.3. Integration within the INGC Infrastructure

In order to support the Pesticides in Groundwater use case within INGC, we had to import a part of the GEUS groundwater database into the INGC platform. The initial import was done in four steps:

1. Create a local PostgreSQL database with the GEUS data-model and a snapshot of the database.
2. Create GEUS as a provider within INGC and getting the DBMS connection info through the INGC API Master Service and Elastic Database Service[4].
3. Upload a dump of the local PostgreSQL database unto the INGC platform through the INGC API Data Import Service[5].
4. Access the instance using Amazon Console, to import the dump into the PostgreSQL database from within INGC[6].

In order for the web GIS client to access the groundwater data, we had to implement web-services to communicate through. We choose to use REST-style services implemented in PHP and returning either JSON data, csv data, HTML code or an image which then could be integrated within the web GIS client. These PHP scripts were deployed on the INGC web server through the INGC API Data Import Service[7].

In order to keep the data within the PostgreSQL database up-to-date, we have created a synchronization scheme which utilizes delta updates. The first part is a locally hosted web-service, which exposes the records required for synchronization. The second part is a provider specific program running at INGC, which fetches the records from the web-service and performs the necessary database operations. The second part was uploaded to the INGC platform using the INGC API Data Import Service[8] and is scheduled using the INGC Scheduling Service[9].

[4]http://ingeoclouds-api.isti.cnr.it/
[5]http://ingeoclouds-api.isti.cnr.it/
[6]Link to Amazon Console or documentation for accessing instances.
[7]http://ingeoclouds-api.isti.cnr.it/
[8]http://ingeoclouds-api.isti.cnr.it/
[9]http://ingeoclouds-api.isti.cnr.it/

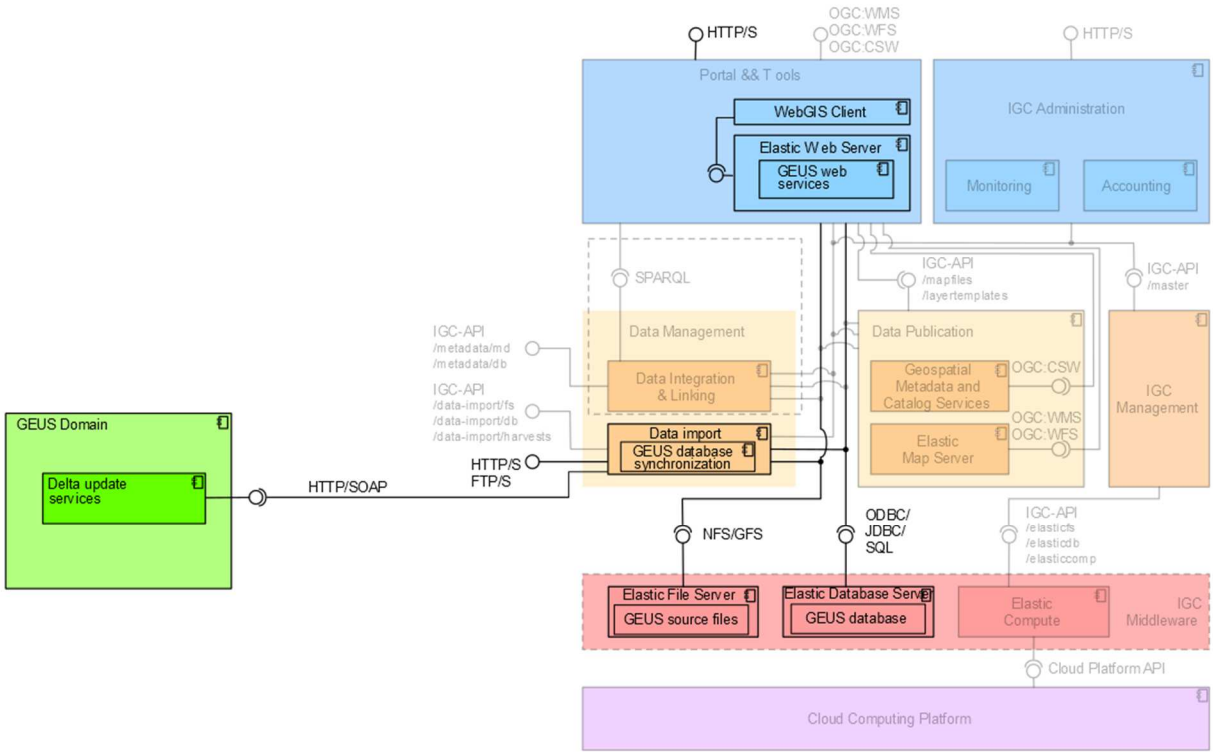


Figure 5: Services for Pesticides in Groundwater Use Case in the INGC system

Figure 5 shows where the different services supporting the Pesticides in Groundwater Use Case are located in the INGC system and which interfaces they use. The services are grouped into the following:

- GEUS web services: These are data services exposed to the web GIS client. The web GIS client can fetch borehole data and other data as described above. The services are written on PHP and use direct JDBC connections to query data within the PostgreSQL database.
- GEUS data synchronization: This program is responsible for updating the database within INGC. GEUS has exposed a locally hosted web service (the Delta update service) from where this program can fetch delta update records for the database within INGC. The harvest module executes this program daily on workdays. The program use direct JDBC connections to access the data within the PostgreSQL database.
- GEUS source files: This is the storage for the java executable file, java log files and the PHP script files.
- GEUS database: Here resides the PostgreSQL database containing the GEUS data on the INGC system.
- Delta update service: This service is responsible for exposing records of new, updates and deleted data in the local database at GEUS.

2.2.4.4. Influence on the Design of the INGC Resulting from the Use Case Implementation

In the original setup of INGC, each provider had only one user created within the PostgreSQL database. Best-practice states, that you need one user to be the data owner and one or more application-specific users with only the absolute necessary rights to access the data. This is done for security reasons, as application-users are written either in clear text configuration files or hardcoded within the application (easily reversed engineered). If the application-user is leaked you do not want it to be the data owner with full rights. Another positive side-effect, is the ability to audit which application does what and when within

the database. If one application executes unwanted or excessive behavior within the database, the audit-log can easily show which application it is.
This led to the development of as an API function – within the elastic database API – which can create additional database users. Once the user is created, the data owner can grant the necessary rights within PostgreSQL.

This use case has a need to be able to upload executable scripts into the INGC platform and have them scheduled for execution. This is needed for harvesting services. A proposal was made, to use the data import API to upload the executable into the provider bucket within the INGC platform and a new API function to schedule the execution using cron-style syntax.

2.2.4.5. Concluding Remarks on InGeoCloudS Integration Work

The most significant novelty introduced in the use case under consideration requires the incremental database synchronization. This method is not necessarily needed in an in-premises deployment. On the other hand this allows having fast update since only the modified data is transferred over the network, and it makes it possible for the data provider to prepare locally a view/sample of the data he wants to make publicly available.

Overall the amount of integration work can be considered “fair” because of the choice of standard (and well known) components within INGC, secured APIs and providers’ workspaces and the direct and easy access to these components.

2.2.5. GROUND WATER RESOURCES MANAGEMENT IN GRANULAR AQUIFERS (UC5 – WAS DOW/UC4)

The use case provides data from both field measurements (groundwater table fluctuation, pumping tests and pollution sources) and chemical analyses (major ions and trace elements). A database has been developed which also consists of various geospatial data. Considering all the above data and the scientific knowledge of geological, hydrological, hydro-geological and hydro-chemical characteristics and properties of the study areas, important interdisciplinary and multi-layered conclusions could be conducted such as: Water balance estimation and assessment, Piezometric surface maps for dry and wet periods’ piezo maps, Hydro chemical maps.
We will provide chemical analysis for major elements and hydro chemical diagrams, charts and graphs pies (images) as map services: Salinization maps (Seawater intrusion), Groundwater pollution and contamination [Inventory of contamination points sources, Regional contamination (F, NO_3, Cl etc)], Trends and statistical analysis of major ions and trace elements.

The groundwater use case was not implemented within the EKBAA IT infrastructure before the InGeoCloudS project.

2.2.5.1. Mapping from Use Case Requirements to Technical Requirements

#rq	Description	Evaluation
1-9	1. Publish maps on the web 2. Facilitate the re-use of my data 3. Dissemination of the contour maps of the aquifers 4. Dissemination of the water table	Web map available in the INGC for my information

#rq	Description	Evaluation
	contour maps 5. Dissemination of the water table and piezometric maps 6. Dissemination of the hydraulic conductivity and transmissivity values plots 7. Dissemination of the iso-chemical maps of various chemical compounds Dissemination of the salinization maps Dissemination of the groundwater pollution - contamination maps	
	Resulting technical requirements and technical choices	
	None specific yet.	

#rq	Description	Evaluation
10	Facility of re-use for: contour maps, water table contour maps, water table and piezometric maps, hydraulic conductivity and transmissivity values plots, iso-chemical maps of various chemical compounds, salinization maps and groundwater pollution/contamination maps without use of support team	Web service to re-use the maps as WMS & download
	Resulting technical requirements and technical choices	
	None specific yet.	

#rq	Description	Evaluation
11	Every day availability of the: contour maps, water table contour maps, water table and piezometric maps, hydraulic conductivity and transmissivity values plots, iso-chemical maps of various chemical compounds, salinization maps and groundwater pollution/contamination maps in INGC system	Daily availability of maps
	Resulting technical requirements and technical choices	
	None specific yet.	

#rq	Description	Evaluation
12	Web-application must contain a map with the following functionalities: 1. Zoom in and out.	Functions are available

#rq	Description	Evaluation
	2. Pan. 3. Display topographical background map. 4. Display other background maps of user's choice including Corine Land Cover and OneGeologyEurope surface geology. 5. Enable user to select and display other maps available as WMS/WFS. 6. Display of the: contour maps, water table contour maps, water table and piezometric maps, hydraulic conductivity and transmissivity values plots, iso-chemical maps of various chemical compounds, salinization maps and groundwater pollution/contamination maps & their legends	
	Resulting technical requirements and technical choices	
	A WebGIS client application in the InGeoCloudS framework has to be built with the ability to fetch boreholes through queries using the Data Linking component Description of functionalities implementation (see section 2.2.5.2) and display their vector data on a map. Furthermore, the user should be given the ability a) to call kriging WPS in order to produce contour maps, water table contour maps etc. a InGeoCloudS infrastructure and b) to fetch and display other maps available as WMS/WFS through the Elastic Map Server and the Geospatial Metadata and Catalog Services components.	
#rq	Description	Evaluation
13-16	Contour maps, water table contour maps, water table and piezometric maps, hydraulic conductivity and transmissivity values plots, iso-chemical maps of various chemical compounds, salinization maps and groundwater pollution/contamination maps available at Regional level, available for the Planners, policy – decision makers and stakeholders level, for the General public/citizens interested in groundwater quality and quantity and for the Large & small Groundwater Abstractors	Accurate & reliable groundwater resources information
	Resulting technical requirements and technical choices	
	None specific yet.	

2.2.5.2. Description of functionalities implementation

Function	Description	Documentation

Fetch borehole data	Fetch from Virtuoso triple-store a collection of borehole objects adhering to filter criteria	In order to support this functionality, we implemented a web-service which takes all the filter criteria as well as the bounding box of the viewport as input parameters and returns GeoJSON data. For each borehole within the bounding box it filters using the different criteria and returns the name, unique identifier and location of the resulting boreholes.
Render borehole data	Draw positions of fetched boreholes on map	The Web GIS client uses OpenLayers for rendering the maps. OpenLayers have built in functionality to render vector layers with data coming from GeoJSON services. This draws an icon on the map using the location from the GeoJSON list. The web GIS client is appending the filter criteria to the request performed by OpenLayers to the borehole service described above.
Display list of boreholes	Display a list of boreholes	Implementation of an array in a panel (WEB GIS client) for displaying the attributes of the selected boreholes.
Display borehole details	Display all relevant information for a borehole including use, topology, time series of concentrations of chemical compounds.	In order to support this functionality, we implement a web-service which applies a SPARQL query over Virtuoso Store and returns the borehole from which the sample was taken as well, the chemical compounds along with their concentrations which are [>, <, =] the maximum accepted limits. Candidate compounds are Ca, Mg, Na, K, HCO_3, Cl, SO_4, NO_3, F.
Download measurements for a borehole	Allow to download data included in the portal	In order to support this functionality, we implement a service which takes a list of borehole identifiers as input parameter and returns a zip-file containing all relevant data as csv-files from the groundwater database related to these boreholes. Each csv-file corresponds to a single table within the data model.

2.2.5.3. Integration within the INGC Infrastructure

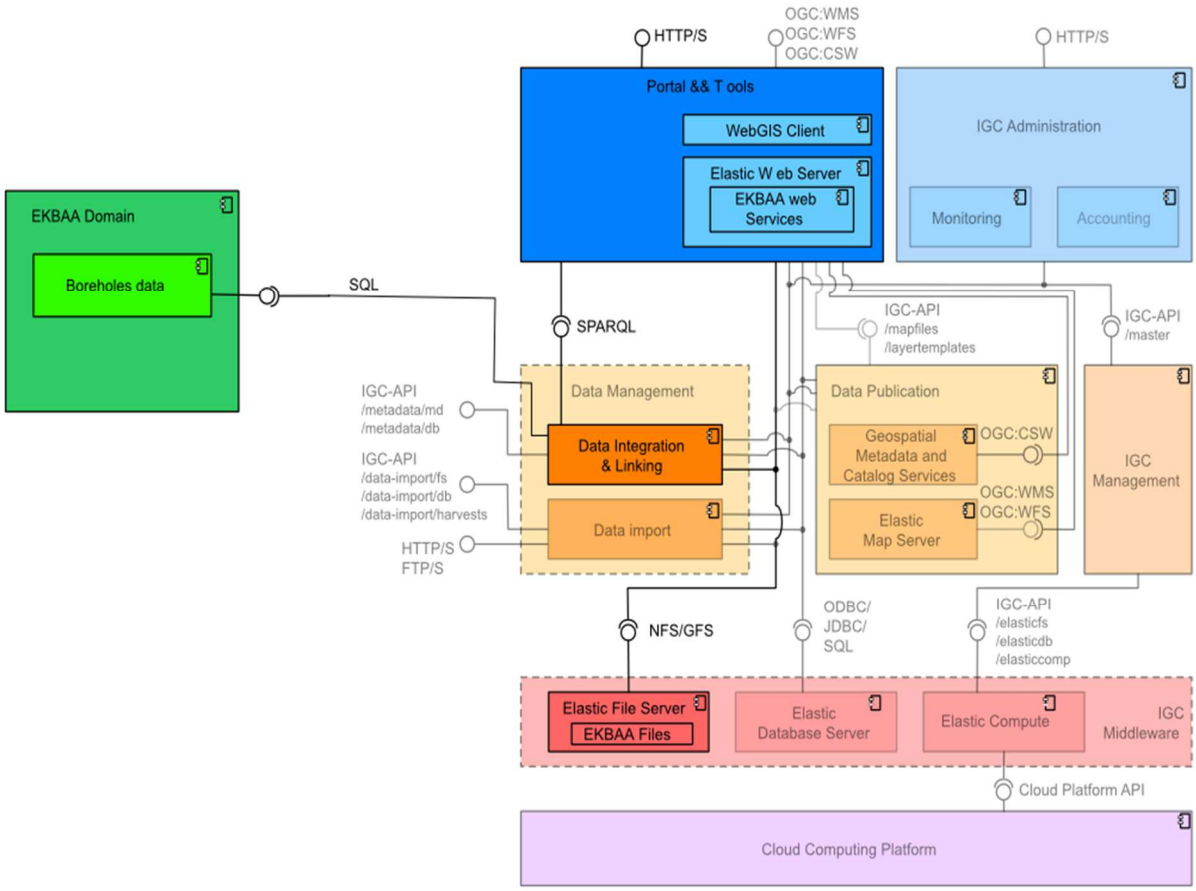


Figure 6 Services for the Ground Water Resources Management in Granular Aquifers Use Case in the INGC system

Figure 6 shows where the different services supporting the groundwater use case of EKBAA are located in the INGC system and which interfaces they use. The services are the following:

- EKBAA web services: These are data services exposed to the web GIS client. The web GIS client fetches borehole data as described above. Data is planned to be retrieved from the Triple Store using the method *ldquery* of the LinkedDataServiceImpl API (D3.2) and the result of the method is transformed (if necessary) to the appropriate geo-format (e.g. GeoJSON) using the method geoldtransform of the GeoLinkedDataServiceImpl API (D3.2).
- EKBAA files: This is the storage that should be reserved for (source, executable etc.) implementation files, if such a need arises for the development of the use case.

Furthermore, in the same figure are shown the components involved to the importing of a part of EKBAA 's groundwater PostgreSQL data to Virtuoso TripleStore. Data import has been performed by FORTH in three steps:

- First, Virtuoso was connected to the relational data indirectly by copying the relational data from that source.
- Secondly, the concepts and relationships contained in the relational data were mapped into corresponding concepts and relationships from the Geo-Scientific Observation Model (D2.2) along

with explanations over the mappings that were performed from notions of EKBAA 's relational data to those of the model.

- Thirdly and after have created the mappings, they were formalized using the R2RML language[10] and registered in the TripleStore using the method addR2RMLMappings from the LinkedDataServiceImpl API.

2.2.5.4. Influence on the design of the INGC resulting from the Use Case Implementation

None.

2.2.5.5. Concluding Remarks On InGeoCloudS Integration Work

Integration is not fully completed at time of writing.

2.2.6. LANDSLIDE USE CASE (ACTIVE LANDSLIDE INVENTORY MAPPING AND SUSCEPTIBILITY ZONING) (UC6 – NEW)

The use case will provide an active inventory map of the occurred landslides updated after every new event recorded in the database. It will be possible to retrieve data concerning the landslides' characteristics (type of movement, causes, season and year of occurrence,..), as well as any information available for the region of occurrence (geology, precipitation, altitude, slope,…). Taking into account the above data, a division of the land into homogeneous areas will be produced according to the landslides' occurrence percentage (Landslides' density map). Additionally, it will be possible for the users to select one or more specific parameters or characteristics and calculate the frequency of landslides accomplishing these criteria. More specific, users will be able to select landslides by location or landslides by characteristics such as their type, the season - year of their occurrence, their dimensions or the type of the area of occurrence such as lithology, slope inclination, altitude, etc.

The calculation of the spatial probability of landslide occurrence will produce a susceptibility zoning map available to the system. The map will be a result of the analysis between the spatial distribution of the landslides (landslides' density) and a group of generative causes (geological, topographical, hydrological etc characteristics of the area) based on the fact that landslides in the future will occur under the same circumstances that they occurred in the past. The calculation of the density map will be redone every time new data for the study area will be provided to the database.

2.2.6.1. Mapping from Use Case Requirements to Technical Requirements

#rq	Description	Evaluation
1-5	Publish maps on the web Facilitate the re-use of my data Dissemination of the Inventory maps of the existing landslides Dissemination of the Density maps of the exposed landslide events Dissemination of the Susceptibility maps revealing areas of potential landslide	Web map available in the INGC

[10] http://www.w3.org/TR/r2rml/

#rq	Description	Evaluation
	occurrence	
	Resulting technical requirements and technical choices	
	None specific yet.	
#rq	Description	Evaluation
6	Have a catalogue of data and services compliant with INSPIRE requirements (regulation)	Catalogue of data and services available
	Resulting technical requirements and technical choices	
#rq	Description	Evaluation
7	Facility to re-use the Inventory, density and susceptibility maps without calling on the support team	Web service to re-use the maps as WMS & download
	Resulting technical requirements and technical choices	
	None specific yet.	
#rq	Description	Evaluation
8	Visualisation and download services as required by INSPIRE and propose harmonized dataset defined by INSPIRE	INSPIRE compliant web services and INSPIRE data model
	Resulting technical requirements and technical choices	
	None specific yet.	
#rq	Description	Evaluation
9	Inventory, density and susceptibility maps available every day in INGC system	Every day availability of maps
	Resulting technical requirements and technical choices	
	None specific yet.	
#rq	Description	Evaluation
10	Web-application must contain a map with the following functionalities: 1. Zoom in and out. 2. Pan. 3. Display topographical background map. 4. Display other background maps of user's choice including Corine Land Cover and OneGeologyEurope surface geology. 5. Enable user to select and display other maps available as WMS/WFS. 6. Display of Inventory map & its legend	.Functions available

#rq	Description	Evaluation
	7. Display of density map & its legend 8. Display of susceptibility map & its legend 9. Show information when location is clicked 10. Show information about landslide when an event is clicked	
	Resulting technical requirements and technical choices	
	A WebGIS Client application in the InGeoCloudS framework has to be built with the ability to fetch landslides vector data through queries using the Data Linking component (see table in section 2.2.6.2 below) and display them on a map. Furthermore, the user should be given the ability a) to call kriging WPS in order to produce inventory and susceptibility maps in the InGeoCloudS infrastructure and b) to fetch and display other maps available as WMS/WFS through the Elastic Map Server and the Geospatial Metadata and Catalog Services components. To produce density maps a Density Kernel OGC-WPS should be implemented in the InGeoClouds framework, having as input the landslides vector data.	
#rq	Description	Evaluation
11-15	Inventory, density and susceptibility maps available for general public, Land use planners, Insurance companies & risk managers (civil protection agencies at national or/and local level) and building companies	The map is available every day
	Resulting technical requirements and technical choices	
	None specific yet.	

2.2.6.2. Description of Functionalities Implementation

Function	Description	Documentation
Fetch landslide data	Fetch from Virtuoso triple-store a collection of landslide events adhering to filter criteria	In order to support this functionality, we will implement a web-service which takes all the filter criteria as well as the bounding box of the viewport as input parameters and returns GeoJSON data. For each landslide within the bounding box it filters using the different criteria and returns the name, unique identifier and location of the resulting landslide.
Render landslide data	Draw positions of fetched landslides on map	The Web GIS client uses OpenLayers for rendering the maps. OpenLayers have built in functionality to render vector layers with data coming from GeoJSON services. This draws an icon on the map using the location from the GeoJSON list. The web GIS client is appending the filter criteria to the request performed by OpenLayers to the landslide service described above.

Display list of landslides	Display a list of landslides.	Implementation of an array in a panel (WEB GIS client) for displaying the attributes of the selected landslides.
Display landslide details	Display all relevant information for a landslide including movement type, altitude, sliding depth etc.	In order to support this functionality, we will implement a web-service which applies a SPARQL query over Virtuoso Store and returns the landslide events along with their corresponding details.
Enable user to select list of landslide events to download	Allow to download data included in the portal	In order to support this functionality, we implemented a service which takes a list of landslide identifiers as input parameter and returns a zip-file containing all relevant data as csv-files from the landslides database related to these landslides. Each csv-file corresponds to a single table within the data model.

2.2.6.3. Integration within the INGC Infrastructure

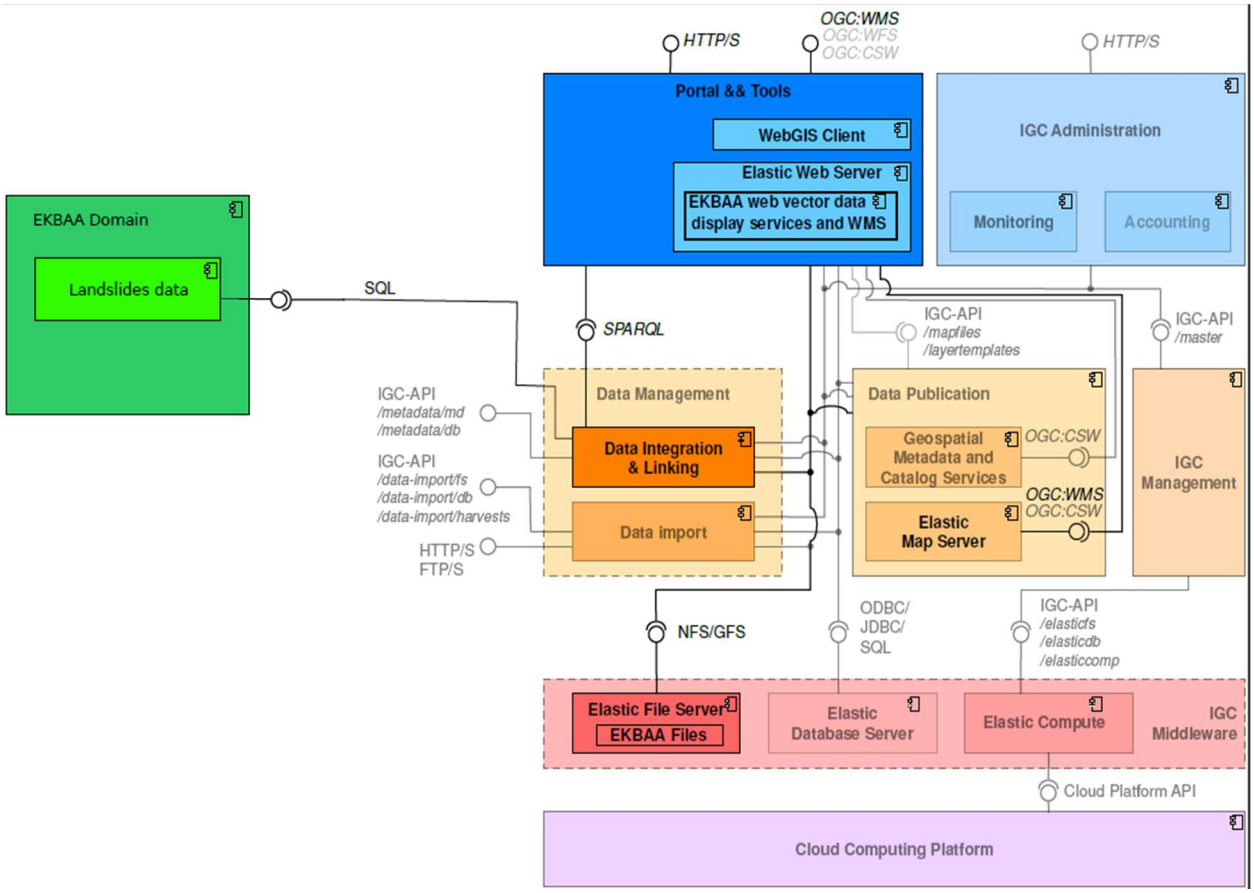


Figure 7 Services for the Landslide Use Case in the INGC system

Figure 7 shows where the different services supporting the landslides use case of EKBAA are located in the INGC system and which interfaces they use. The services are the following:

- EKBAA web vector data display services and WMS: These are data services exposed to the WebGIS client. The WebGIS client fetches landslides data as described above. Data is planned to be retrieved from the Triple Store using the method *ldquery* of the LinkedDataServiceImpl API (D3.2) and the result of the method will be transformed (if necessary) to the appropriate geo-format (e.g. GeoJSON) using the method *geoldtransform* of the GeoLinkedDataServiceImpl API (D3.2).

WMS service is exposed to the WebGIS client to display EKBAA's density, inventory and susceptibility maps and to others who want to download the maps as WMS to use them in other applications.

- EKBAA files: This is the storage that should be reserved for (source, executable etc.) implementation files, if such a need arise from the development of the use case. In addition, the storage is used to store files that will become available as the output of kriging WPS, in grid data format. In order for the WebGIS Client to access the gridded data, a service has to be implemented (with map server) which will create WMS services from our files.

Finally, in the same figure are shown the components involved to the importing of a part of EKBAA 's landslides PostgreSQL data to Virtuoso TripleStore. Data import was handled by FORTH in three steps:

- First, Virtuoso was connected to the relational data indirectly by copying the relational data from that source.
- Secondly, the concepts and relationships contained in the relational data were mapped into corresponding concepts and relationships from the Geo-Scientific Observation Model (D2.2) along with explanations over the mappings that were performed from notions of EKBAA 's relational data to those of the model.
- Thirdly and after have created the mappings, they were formalized using the R2RML language and registered in the TripleStore using the method addR2RMLMappings from the LinkedDataServiceImpl API.

2.2.6.4. Influence on the design of the INGC resulting from the Use Case Implementation

None.

2.2.6.5. Concluding Remarks On InGeoCloudS Integration Work

Integration is not fully completed at time of writing.

2.3. GEO PROCESSING SERVICE

Geoprocessing refers to ordinary kriging interpolation and is implemented as a WPS service within the InGeoCloudS project. The service was coded using the open source 52 North WPS[11] java implementation (deployed by tomcat server), which follows the last OGC-WPS 1.0.0 standard[12]. According to the standard, input/output of the WPS are in general described by XML documents. Kriging is performed on a set of points in the form of [x, y, value] format (input data) and requires a number of parameters for fine-tuning the interpolation process. The result is a grid filled with a predicted value per grid point. 52 North WPS implementation supports vector data input in a number of formats, including ESRI-Shapefile[13](zipped) and GML2.0/3.0[14]. Those data are embodied in the input XML which is posted to the service. Input XML includes, in further, kriging parameters. In current implementation the resulting XML consists of temporary links to a) a file with the input data in csv format, b) the predicted values in csv format and c) the preview of kriging result as a png image. Internally, kriging is performed using the open source library gstat[15]for R. Two of its main advantages are that gstat implementation is fast enough (even if it is executed in the framework of R) and has the ability to perform kriging on unprojected data and thus no re-projection is needed. The interconnection between java code and R is provided by the Rserve[16]TCP/IP server.
A client WPS web application is now built which is based on OpenLayers[17]and gives the user the ability to a) receive input data, b) provide kriging parameters, c) perform kriging on input data by calling the OGC-WPS and d) download through temporary links the output files of the service.

2.3.1.1. Mapping From User Requirements to Technical Requirements

#rq	Description	Evaluation
1	OGC-WPS requirements: 1. INPUT: data in [x, y, value] format. The WPS service will accept input data in one of the vector data formats supported by 52 North WPS. It is assumed that CRS of data is included in input data, otherwise should be provided as one more input argument. 2. PARAMETERS: • Attribute Field of input data (denoted as “value” in vector data [x,y,value]) used for kriging interpolation. • Ordinary kriging parameters: Range, Nugget, Sill and Semivariance model (i.e. spherical, exponential etc) • Local neighbourhood prediction: SearchRadius, MinNum, MaxNum In most kriging implementations, a subset of points with known values is defined in an area of an acceptable, user-	OGC-WPS implemented but not yet available in INGC

[11] http://52north.org/communities/geoprocessing/wps
[12] http://www.opengeospatial.org/standards/wps
[13] http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
[14] http://www.opengeospatial.org/standards/gml
[15] http://cran.r-project.org/web/packages/gstat/index.html
[16] http://www.rforge.net/Rserve
[17] http://openlayers.org

#rq	Description	Evaluation
	given radius (SearchRadius) around the point in question and only this subject is used for prediction (local neighbourhood prediction). Furthermore, if the cardinality of this subset is less than a user-given value MinNum, no prediction is made and if exceeds a user-given value MaxNum, then only MaxNum points will be used in prediction. • Gridded data cell size (grid extent will be automatically computed by the input data). 3. OUTPUT: Temporary links of a) the initial and the gridded data in txt, csv or GeoJSON format and b) the preview of the kriging result as common WMS layer or alternatively as an image having jpeg, png etc format. 4. Data processing speed: the computational time of kriging depends on the size of the output grid and the number of points with known values that are used for each prediction. However, the ability to use local neighbourhood prediction greatly reduces (in most cases) the processing cost per interpolated point, without decreasing the accuracy of computations. Typical processing time should be about 3 sec. for 100.000 predictions.	
	Resulting technical requirements and technical choices	
	The requirements mentioned above are already fulfilled by the OGC-WPS service that has been implemented. Preview is returned as png image. The implementation of the service in the InGeoCloudS infrastructure will affect the Elastic Web Server component.	

#rq	Description	Evaluation
2	Client web-application with the following requirements: 5. Ability to take input data from a user, independently or additionally to pre-existing data-sets, in order to use them as input data for the kriging procedure. 6. Implementation of a form with items that correspond to the parameters of Kriging procedure in order to give the user the ability to fine tune the provided service. 7. Ability to forward the input data as well as the parameters to the OGC-WPS service and to parse the response xml of the service. 8. Downloading of the initially used data and the results of kriging through temporary download links.	Client web-application not yet completed
	Resulting technical requirements and technical choices	
	An input form integrated in the general purpose Web GIS Interface that will be implemented for the ground water use case, has to be built in order to fulfil this requirement. The form will accept input data from the currently loaded vector layers on the map or from a selection of the currently loaded layers. In the InGeoCloudS framework the necessary input data manipulation (e.g. conversion from Geojson to GML2 format) will be carried out in the OpenLayers level of the WebGIS Client.	

#rq	Description	Evaluation
	The use of temporary download links (instead of storing the input/results in INGC db) is justified by the complexity of kriging procedure which may lead a user to create many, temporal in nature, results, that otherwise would have been stored in db. As an alternative scenario, the user could be prompted to save the temporary links in INGS db. Considering the InGeoCloudS infrastructure, the component affected is the EKBAA/IGMEM workspace (provisioned by the Elastic File System module) used to store the temporary results available for downloading.	

2.3.1.2. Integration within the INGC Infrastructure

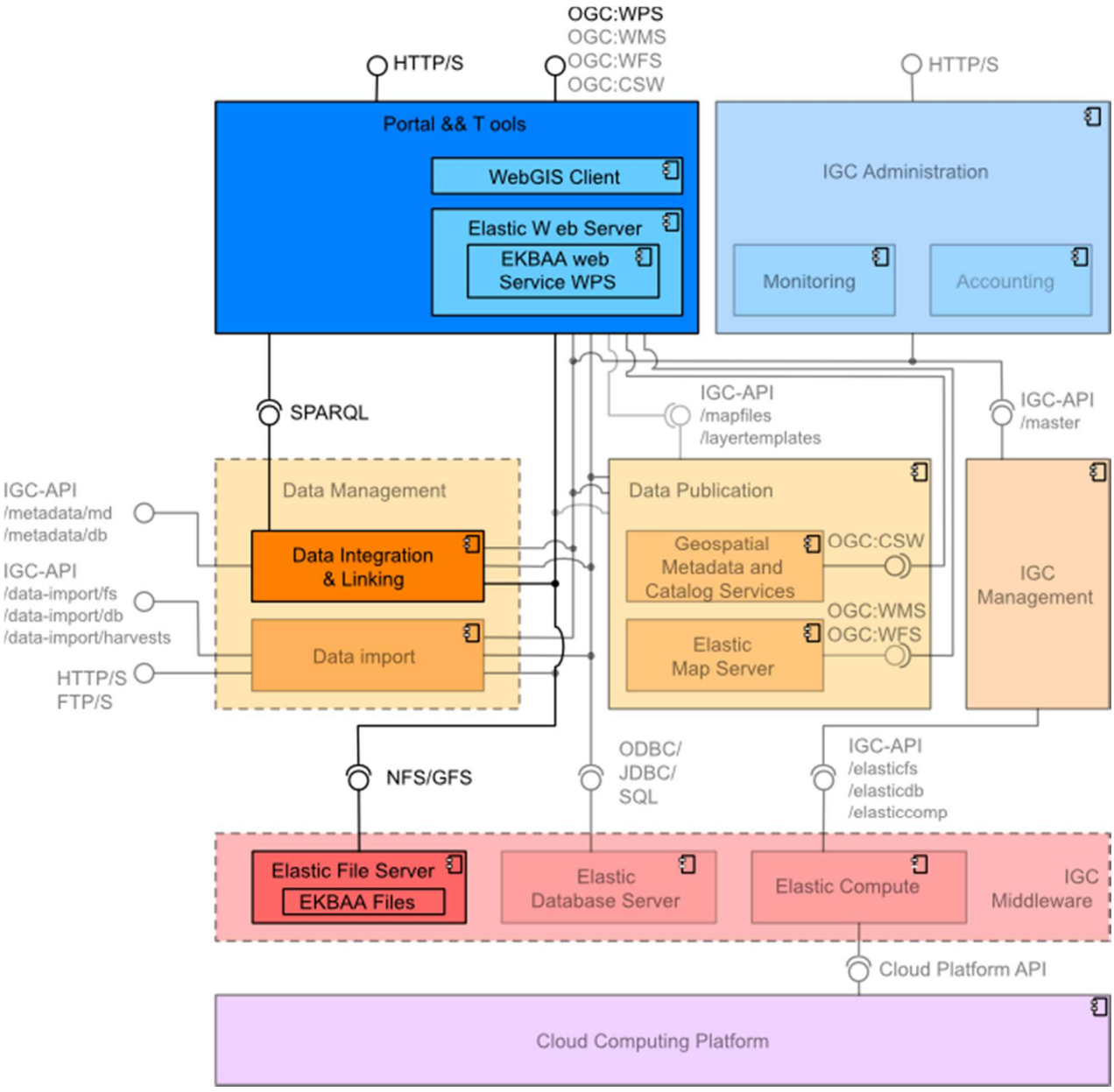


Figure 8: INGC infrastructure used by the Kriging Geoprocessing Service Use Case

As it is shown in Figure 8, the kriging WPS is located in the Elastic Web Server component of the diagram. Input data are fetched from the Virtuoso Triple Store using SPARQL queries through the *ldquery* method of the LinkedDataServiceImpl API (D3.2) and the result of the method are transformed (if necessary) to the appropriate geo-format (e.g. GeoJSON) using the method geoldtransform of the GeoLinkedDataServiceImpl API (D3.2).
Retrieved data are displayed by new methods implemented in the WebGIS Client. That client also contains a panel which gives the ability to the user to provide kriging parameters. Kriging OGC:WPS is a service which is meant to be used for interpolating input data of other data providers apart from EKBAA and therefore, its implementation is included in the so called Default WebGIS Client. In addition, the OGC:WPS service would be used by other, external clients.
Finally, the box entitled EKBAA files in the component diagram represents the storage that should be reserved for (source, executable etc.) implementation files if needed from the development or usage of the WPS. However, it is noted that data providers will have the ability to store the resulting files of the WPS to their storage in the Elastic File Server.

2.3.1.3. Influence on the design of the INGC resulting from the Use Case Implementation

None.

2.3.1.4. Concluding Remarks On InGeoCloudS Integration Work

Integration is not fully completed at time of writing.

2.4. LINKED OPEN DATA

The Linked Data Management service can be exploited for performing all sorts of management functions on LD that correspond to the data providers' data sets, such as performing SPARQL queries, importing LD, exporting LD and updating LD. However, it also provides some added-value functions that cannot be easily provided through a provider-specific data management scheme supported by the relational-based (data) management APIs of the INGC infrastructure, such as performing cross-provider or cross thematic queries, apart from data provider-specific ones, and exporting (cross- or provider-specific) LD in a XML-based INSPIRE-compliant form or in different feature collection representation formats. In fact, the latter ability can be integrated with the data publication services so as to be able to offer not only INSPIRE-compliant services but also INSPIRE-compliant data.

The above functionality can be exploited by:

- *Application providers* which desire to: (a) perform cross-provider or cross-thematic queries or even provider-specific queries without digging into the specifics of each provider's data model as well as (b) export LD in XML-based INSPIRE compliant form or different feature collection representation formats
- *Data providers* which: (a) desire to exploit the above functionalities for application providers, (b) work with their existing data sets in their premises without requiring to load them in the INGCS platform, (c) do not desire to implement synchronization code such that the data stored in the INGCS platform are correctly and immediately updated, (d) would like to have their data integrated in such a way that they can be exploited in a fair and uniform manner, and (e) (advanced – would need effort to implement) perform some type of reasoning as well as integrate their data with external data sets (e.g., connect landslide data with temperature data or earthquake data with municipal data such that the cities close to the earthquake are quickly discovered so as to inform accordingly rescue teams as well as civilians, if needed).

A summary of the methods offered by the Linked Management service is provided below, while more detailed information along with some intuitive diagrams which show the interaction between the user/developer and the API/service can be found in the Deliverable *D3.2. Cloud architecture, configuration and data access implementation* as well as in the API's online documentation:

- **/linkeddata/ld/addR2RMLMappings**: Allows the definition of mappings from relational data to linked data conforming to the InGeoCloudS GSOM model.
- **/linkeddata/ld/addXSLMappings**: Allows the definition of mappings from xml data to linked data conforming to the InGeoCloudS GSOM model.
- **/linkeddata/ld/import_status**: Allows the requester to enquire the current status of a non-blocking linked data importing request.
- **/linkeddata/ld/inspire_export**: Allows to export all data related to a specific theme, possibly concerning a specific location, in an INSPIRE-compliant XML form
- **/linkeddata/ld/ldexport**: Allows the exporting linked data statements for particular resources or all linked data content of particular graphs.
- **/linkeddata/ld/ldimport**: Allows the importing of linked data to the underlying triplestore. The linked data to be imported are either stated in the HTTP body or found in a specific URL given as input. Obviously, the first option is more appropriate for linked data files with a small size as the client will wait for a long time until it obtains the result.
- **/linkeddata/ld/ldquery**: Allows the issuing of SPARQL queries with optional values for the timeout and maximum rows parameters.
- **/linkeddata/ld/ldupdate**: Allows the updating of linked data through SPARUL statements.
- **/linkeddata/ld/transform**: Allows transforming an RDF file from one format to the other.

- **/linkeddata/geold/geoldquery**: Allows issuing GeoSPARQL queries with optional values for the timeout and maximum rows parameters. In this way, the application/data providers cannot only perform normal SPARQL queries but also geo-spatial ones, thus finally reaching the functional power of relational + PostGIS solutions.
- **/linkeddata/geold/geoldtransform**: Allows transforming geo-spatial SPARQL results to feature collection representations in particular formats (KML, GeoJSON, Shape, GML).

Based on its above added-value, advanced querying and exporting as well as LD management functionality, the Linked Data Management service can be regarded as either a complementary or an alternative mechanism for supporting any use case and managing the respective use-case data with respect to the provider-specific data management mechanisms that are available in the INGCS platform.

We foresee two different scenarios with increased exploitation of the service's functional capabilities irrespectively of the (LD) importing mode exploited for generic applications/use cases. In particular, the first case, also depicted in Figure 5, shows how the service can be used to enhance the querying capabilities of the use cases/applications, without requiring to modify the rest (functionality/code) of the application. The second case, depicted in Figure 6, shows how the service can be exploited to provide complete support to the application/scenarios through the management of the corresponding (linked) data, the issuing of the respective queries and the exportation of the appropriate information. It should be noted that in both cases, the application can still operate on the original data (e.g., insert or modify existing data) but have the respective LD also updated either manually (by generating LD specifications and corresponding SPARUL updates) or through the automatic R2RML mapping mechanism.

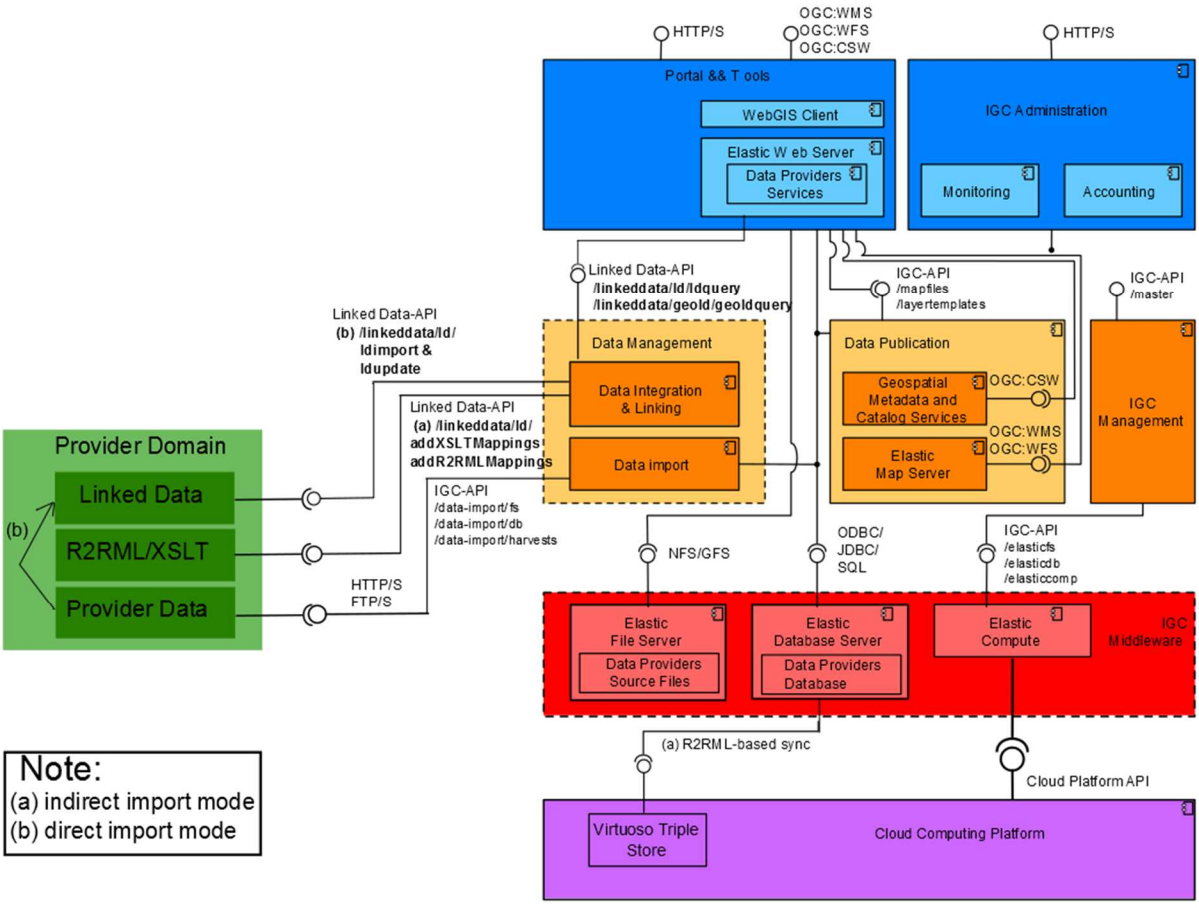


Figure 9. The first exploitation case of the Linked Data Management Service

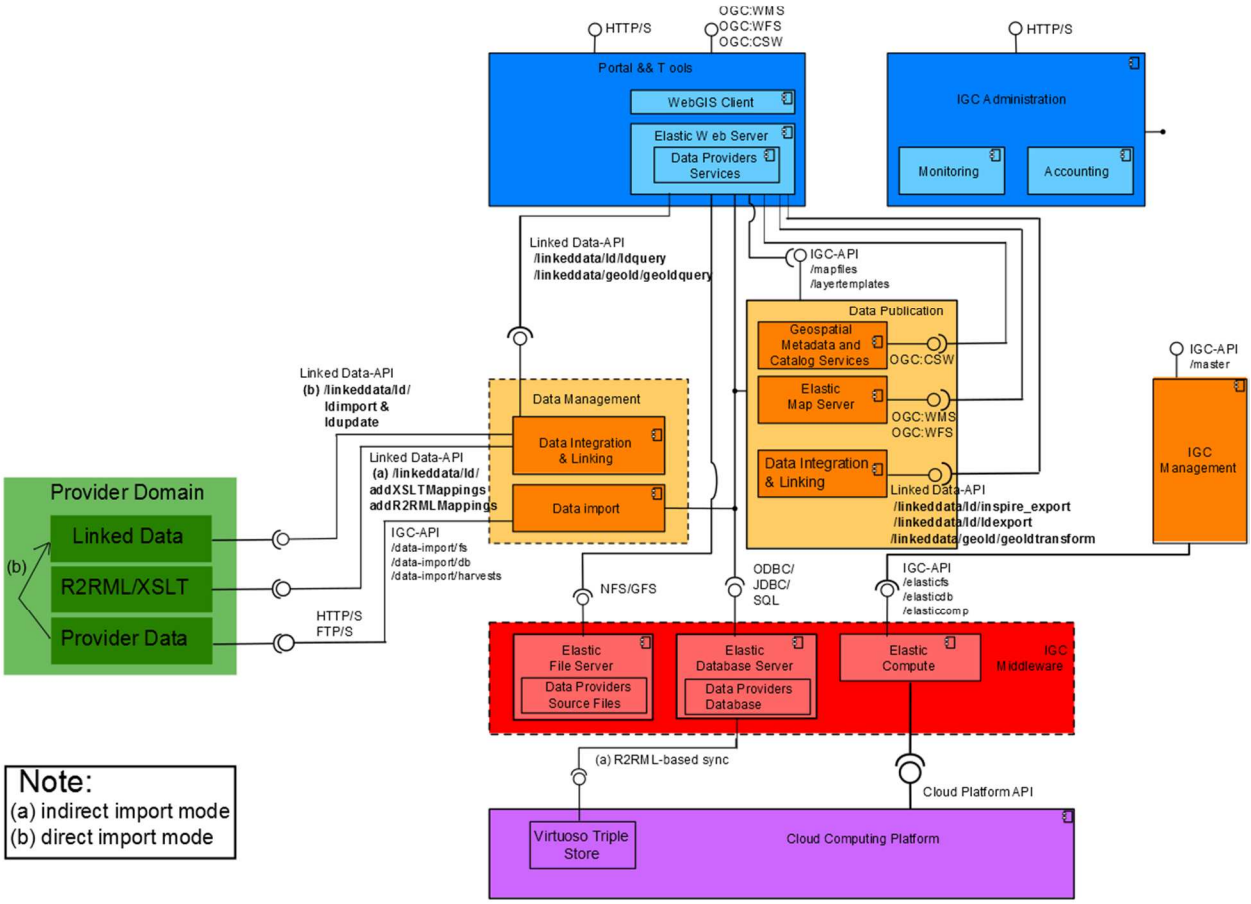


Figure 10. The second exploitation case of the Linked Data Management Service

Based on the above rationale, any of the use-cases designated in the project can benefit from the functions provided by the Linked Data Management service with little or small adjustments on the implementation code. In fact, apart from the integration work required for e.g. providing INSPIRE-compliant services and data, the main modifications to be performed are the following:

- Change the way the provider's data are imported: Here, as already described, two importing modes are supported: direct and indirect. In the direct mode, the data provider is in full control of the import procedure, where he/she should transform his/her data specifications in the original form into a LD form, obviously through establishing a mapping from his/her data model to the GSOM model. In this way, the data provider will import his/her LD through the respective function of the Linked Data Management API as well as guarantee that any updates to his/her (original) data in the initial form are mapped to respective SPARUL statements to be used as input to the respective method of the API.
- In the indirect mode, the data provider, depending on the form of his/her original data, should be able to produce a mapping to the GSOM according to the R2RML or XSLT standard specifications, respectively, and establish it by calling the respective API method. In this way, any updates to the original, relational data will be able to immediately and automatically produce the respective updates on the respective LD. Please note that in this second case, any updates to the XML-based data of the provider are not automatically enforced but these updates should be fed to respective API method call in order to update the respective LD, accordingly.
- In any of the two import modes, the data provider should have a good knowledge of LD technologies as well as of the GSOM model. In the direct mode, the data provider should use tools, such as XSLT editors and executors or RDFiers in order to generate LD out of his/her

original data according to a (possibly implicit) mapping between his/her data model and GSOM. In the indirect mode, the data provider should be able to produce R2RML or XSLT specifications depending on the original form of his/her data sets. While there is good support for XSLT, the R2RML can also be supported through exploiting existing LD editors but, of course, this requires having some expertise not only in using these tools but also concerning the specifics of LD specifications.

- Based on our current experience and on the fact that FORTH was responsible for producing the respective mapping specifications as most of the data providers did not have any expertise on using LD technologies, the data providers were not actually burden with the mapping tasks. However, after the end of the project, a new data provider with no LD technology expertise will have a difficulty in providing such mappings. In this case, such a data provider could benefit from additional consulting services (e.g. offered by FORTH) which would provide assistance in the production of the mappings. On the other hand, new data providers with LD expertise can benefit from the existing documentation concerning the GSOM model and the importing functionality of the Linked Data Management service.
- Change from (traditional) SQL to the SPARQL-way of posing queries. This again requires that the data provider has a good knowledge of the LD technology, and especially SPARQL, as well as of the respective part of the GSOM model which corresponds to the provider's data model. As SPARQL resembles a little bit SQL, even data providers with no LD expertise could be able to pose SPARQL queries by studying the existing, widely available documentation of SPARQL and LD (in the Internet) so as to be able to change the way of thinking from records to triples. In any case, FORTH can also provide consulting services for enabling data providers to transform their SQL queries to SPARQL ones. Based on our current experience in the project, where the data provider partners did not have any experience in SPARQL, such a transformation was performed by the FORTH team. But we believe that once the IT parts of the data providers become aware of SPARQL, they could easily perform such a transformation themselves. In any case, it would be also interesting to see any additional developments in the LD field so as to update the respective API methods. For example, once we have a XSLT or R2RML mapping, it would be interesting to see whether an automated transformation from SQL queries to SPARQL queries could be realized and incorporated in the API.
- Use the to-be-integrated exporting INSPIRE-compliant functionality or the exporting functions provided by the Linked Data Management API. This just requires exploiting the respective API method. In case that not all data need to be exported but only a subset of them, then the data provider can use first the SPARQL query functionality to produce the required SPARQL results corresponding to the needed data set and then the export functionality to produce feature collection representations (or even INSPIRE-compliant form if this is implemented). The latter case will again just require good knowledge of SPARQL (or exploitation of the consulting services of FORTH).

We believe that the above three code/query modification steps are not only subtle but also quite easy to realize, especially if we consider that all the appropriate methods are already in place. This is witnessed by the fact that two use cases already exploit some functions of the service (especially the SPARQL query functionality) but other required functionality can also be realized through exploiting other service/API methods, such as those which transform SPARQL results to feature collection representations.

In this way, by also considering that SPARQL results can also be in the csv form, then most of functionality of the case study in e.g. Section 2.5 could be realized through calling functions of the API. The benefits of using the Linked Data Management API have also been acknowledged by another project partner who plans to produce an implementation variant for its case study in Section 2.2.1 (Pesticides in Groundwater Use Case). The API could also be easily exploited by the geo-processing service by providing the

respective input to the algorithm after an input selection step comprising a SPARQL query posing and a transformation of the results to the format required by the implementation software.

Possible GSOM evolutions

As explained in [R3], the process of knowledge elicitation, representation and instantiation is not trivial and requires going through different phases associating experts and knowledge management specialists. Moreover mapping to INSPIRE data models must be maintained (see section 3.1 below for details).

3. CURRENT LIMITATIONS AND FUTURE WORK

3.1. INSPIRE COMPLIANCE FOR GEOSERVICES

The current implementation of the GeoPublication use case and data publication services allow to fulfil an important part of technical requirements published by INSPIRE team: metadata, CSW, WMS, WFS, ATOM, etc. Nevertheless, the current opensource software selected in the infrastructure does not provide a WFS version 2.0 (limitation to the version 1.1) as required by the INSPIRE guidelines. This limitation is not problematic to end-users but the download service – part direct download - is not fully compliant in the current configuration. The development of a WFS version 2.0 is not compatible with the objectives and delays of the INGC project but BRGM's team is in discussion with the opensource community of mapserver to implement a version 2.0 in 2014.
An alternative would be to use another opensource solution but only GeoServer currently proposes a correct implementation of the WFS Version 2.0 (and GML 3.2). But, in fact, the main issue for the compliance to INSPIRE requirements is not the WFS 2.0 implementation but the capability to request (and answer) complex data model (as defined by INSPIRE data model) with the operations provided by the OGC standards. Geoserver and other commercial servers are testing the capability to manage (and configure) the mapping between data model and WFS / GML / Filter Encoding combination – some tests are done by the BRGM in the EU Promine project[18] - but lot of technical challenges (and configuration) are not currently solved and overstep the objectives of the InGeoCloudS project.

The compliance for data model defined by INSPIRE was also a limitation for the INGC infrastructure in Pilot1. The architecture allows to store (in the elastic DB) and publish data compliant with INSPIRE (see description in D2.1). But this means that each data provider should organize his dataset to be compliant with the INSPIRE data model, i.e. map his own data model to the INSPIRE data model, adjust his attributes, use code list defined by INSPIRE, etc.

This work can be done before the integration in the INGC infrastructure but the complexity of the tasks required for this pre-processing postpones the publication of INSPIRE data model (INSPIRE requires this compliance not before 2015 and only for new datasets).

With the Pilot2 and the Linked Data Management Service, an alternative is however the use of the GSOM model.
As GSOM is used for the integration, the data provider is not required to map his data model to the INSPIRE data model, but a mapping between the GSOM and the INSPIRE data model has already been established and realized through the respective export function of the Linked Data Management API. Thus, the data sets exported through the INGCS infrastructure will be INSPIRE-compliant without requiring any additional actions from the data provider apart from either mapping his/her data set to the GSOM via providing R2RML or XSLT specifications or directly importing his/her data set in a LD form (according to the GSOM model).

It is worth noting that the GSOM to INSPIRE mapping is complete (for all thematic fields related to the project) but some data are potentially lost due to INSPIRE model incompleteness (so this is the actual place where we lose information).

InGeoCloudS offers dedicated documentation for facilitating this mapping and as stated in D651 (see [R8]) the consortium also intends to provide consultation services apart from detailed documentation.

[18] http://promine.gtk.fi

Since INSPIRE fosters the use of GML, the development of the relation GSOM2GML and WFS service is planned for further evolutions of the platform in the context of the InGeoCloudS project.

3.2. PERFORMANCE AND OPTIMIZATION FOR GEOSERVICES

The cloud infrastructure and the architecture defined by the INGC project are very smart and offer a scalable solution for the geopublication services. However, the cloud infrastructure remains a computing infrastructure with servers, CPU, memory, disk storage, and software deployed should be optimized to use correctly the infrastructure. The data publication is no exception to the rule and some challenges should be resolved during the project to fulfil all performance requirements:

- an adequate initial server should be used as showed in the following figures.

Figure 11: Comparison of two instances for a charge test of a GetMap operation

- A correct management of the autoscaling system to digest peak of uses of services and to be able to create very quickly and automatically new instances because of the impossibility to foresee the peaks of use.
- The use of cache technology as described in the section 2.1.4.4

3.3. (BIG) DATA UPLOAD IN THE CLOUD

A current limitation of the cloud is the delay to push data in the cloud infrastructure. This general limitation could be very significant in the geospatial domain where volume of data could quickly increase (for example, the OpenStreetMap in Europe is stored in more 2 Go Database).
Cloud Services Providers may offer very specific services to solve the issue of uploading large volumes of data. For instance, Amazon allows shipping, to one of its data centres, a physical disk whose content is then copied to the cloud distributed storage (http://aws.amazon.com/importexport/).
This nevertheless is valid only for one-shot transfers and is not applicable to regular update requirements on large data volumes.

Some strategies have been developed within InGeoCloudS to overcome this constraint for our use cases requirements:

- synchronize only new or updated datasets to avoid a “cancel and replace” approach (used by groundwater Use Case)
- push optimised dataset for the use in the cloud in place of original datasets (used by OpenStreetMap base layer with the importation of images instead of the original database).

4. CONCLUSION

The web services implemented on the INGC platform in order to support the selected use cases were mainly newly developed. It is therefore difficult to assess whether the work required to take advantage of the cloud technology in the chosen form is different from traditional implementations. It is however the opinion of the project partners that the implementation work is very comparable to what they have had to carry out on projects with similar complexity and requirements in terms of user functionality. This is probably due to very modular architecture of the INGC platform.

It has been evident that good documentation and user friendly tutorials are crucial for an easy exploitation of the features of the platform as are stable versions of the middleware components.

In the current phase of the project it is unlikely that all limitations of and difficulties in using a cloud platform has been encountered by the partners as several aspects still need to be evaluated during the remaining phase of the project.