



Implementation of an ICT facility for geophysical and airborne data archiving and sharing



Deliverable number:	D7.3
Work package:	WP7 – 7.4 and 7.5
Intermediate Objective:	IO7.5
Deliverable type:	<input checked="" type="checkbox"/> Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify
Dissemination level:	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Restricted
Estimated delivery (bimester):	B12
Actual delivery date:	B12
Author(s) (Partner-OU):	Ilaria Catapano, Valeria Gianpaolo, Enzo Lapenna
Reviewed by:	ITINERIS Executive Board
Note:	

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
 Funded by EU - Next Generation EU
 Mission 4 “Education and Research” - Component 2: “From research to business” -
 Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Table of contents

1. INTRODUCTION.....	4
2. INFORMATION CONTENT REPOSITORY.....	5
3. WEBGIS PLATFORM.....	10
4. WEB SERVICES FOR DATA ACCESS.....	10

1. INTRODUCTION

The deliverable 7.3 is prepared in the context of the ITINERIS project, within the WP7 that entails the activity 7.4 “Networking the ground-airborne instrumental facilities for land surface monitoring” in charge of CNR-IMAA, and the activity 7.5 “Earth Observation infrastructure based on airborne X/L Band SAR and in-situ geophysical data”, in charge of CNR-IREA.

This deliverable aims at illustrating the ICT facility designed to store and share data acquired from geophysical and airborne technologies and the results achieved by processing these data. The ICT facility is expected to have a positive impact on the collaboration among researchers, practitioners and end-users fostering interdisciplinary project. The ICT facility has been designed to be a valuable tool assuring rapid access to existing data and improving knowledge about the territory, which is flexible enough to accommodate future data growth and technological advancements, with the final goal encouraging innovation opportunities. The ICT facility is implemented to be compliant with the ITINERIS HUB of the Work Package (WP) 2 fulfilling meticulous technical specifications and standards to ensure optimal performance and interoperability, enabling data reuse and collaboration, by FAIR and Open Data principles adoptions.

The ICT facility aims at addressing three main purposes:

- a. creating a data repository for research purposes, useful for scientists that want to test their data processing methodologies by using benchmark on field data;
- b. building a national storage center collecting data and results of on demand geophysical surveys and airborne synthetic aperture radar missions, useful for scientists to study the environment conditions of the Italian territory and monitoring its changes and for practitioners and policymakers to improve their knowledge about the investigated areas and to gather, for each area, information about the performed inspections and their results;
- c. making available an internet interface representing an efficient tool for users to access and download data through standardized Web services and that supports various user roles and access levels.

Key components of the facility are:

1. an intuitive interface allowing users to navigate through datasets, access reports, and visualize data through maps and graphs.
2. a robust data management system that utilizes both relational and non-relational databases for different types of geophysical data.
3. implementation of security measures to protect sensitive data and ensure user authentication.

2. INFORMATION CONTENT REPOSITORY

Data and results populating the ICT facility are collected and achieved by ITINERIS sensors operating on demand at different test sites. At each site, the inspections could be carried out by means of one or more sensors, which are chosen according to the needs, the survey aims, the actual availability of the sensors. In addition, starting from the same kind of data (i.e., data collected by the same sensors) different final results can be provided according to the data processing methodology exploited to process the measured data.

Due to the significant heterogeneity of the information, preliminary activities have regarded the definition of the useful materials for scientists and/or end-users (i.e., practitioners and policymakers) to be considered for each test site and for each technology. These materials involve information content about test sites, goal of the surveys and the employed technologies, geospatial data for webGIS visualization of the area under test, typologies and format of raw data, processed data (intermediate results) and final results.

The available information contents are catalogued in three categories: 1. Material for scientists and end-users; 2. Material for scientists; 3. Material for end-users. The contents of each category are structured as follows.

Material for scientists and end-users

List of the ITINERIS technologies with contact person information, List of the investigated test sites, Details about test site and technologies as specified below.

Test site

General description – Motivation of the survey – Period of the survey (year and month) - List of the adopted technologies – List of the specific employed sensors – Georeferenced data about the area under test and its extension.

Technologies

General description (i.e., working principle of the sensor), Datasheet, Dates and time of employment, Survey modality (in situ by hand survey, in situ by using movable platform survey, drone based survey), Physical measured quantity, output description. Contact person name, affiliation, and email address.

Materials for scientists

For each technology and for each test site

Georeferenced data of the investigated area and its extension, Georeferenced data of the measurement points (if available) or their spatial offset, measurement configuration parameters, description of all the available data files (number, acquisition data, name, structure, format), description of the processing chain, description of the files referred to intermediate and final results and the processing parameters (number, name, structure, format), data files, files storing intermediate and final results, processing parameters files.

Materials for end-users

For each test site

Table of content summarizing the performed surveys, the final results provided by all the employed technologies and their links.

Subsequently to the classification of the information contents, the activities have regarded the standardization of their formats and their storage policy for assuring efficient and secure access and management, while protecting sensitive information.

Generally speaking, the information contents of the ICT facility are identified as:

1. Raw Data (L0): Unprocessed data directly from sensors (e.g., seismic readings, aerial images).
2. Processed Data (L1): Data that has undergone analysis (e.g., calibration, filtering, smoothing).
3. Products (L2): Results of analysis (e.g., interpreted geophysical models, images, zoning maps).
4. Metadata: Information about the data, including collection methods, date/time of acquisition, and sensor specifications.

Raw / processed data / products and metadata are organized and stored according to a structured approach, which aims at enhancing the ability of researchers, practitioners and policymakers to find, use, and share information gathered by means of geophysical and airborne technologies in order to facilitate innovative research in geophysical sciences as well as to support informed decision-making. The adopted structured approach is summarized below.

Raw and processed data and products are organized into specific categories based on the type of geophysical investigation:

- Seismic Data: Information from seismic surveys, including waveforms and interpretations.
- Magnetic Data: Data from magnetic surveys, including total magnetic intensity readings.
- Electrical Resistivity Data: Data from resistivity surveys, presenting resistivity profiles and interpretations.
- Ground Penetrating Radar (GPR) Data: Data collected through GPR, including images and depth profiles.
- Aerial Synthetic Aperture Radar (SAR) Data: Data collected by SAR system mounted on airplane, including focused images and digital elevation maps.
- Aerial Optical Data: data collected by electro-optical sensors operating in different spectral ranges;
- Geological Data: Supporting geological information, such as rock types and soil characteristics.

Each category includes: 1. original data collected from surveys, stored in standardized formats (e.g., CSV, HDF5); 2. data that has undergone analysis or transformation, also stored in standardized formats; 3. visualization outputs, i.e. graphs, maps, and other visual representations generated from the data.

Furthermore, each dataset, in each one of the above categories, includes core metadata elements, which may consist of:

- Title: A descriptive title of the dataset.
- Abstract: A brief summary of the dataset's content and purpose.
- Creator: The individual or organization that created the dataset.

- Date of Creation: The date when the data was collected or processed.
- Spatial Coverage: Geographic area covered by the data, including coordinates.
- Temporal Coverage: Time period during which the data was collected.
- Data Quality Information: Information on the accuracy and reliability of the data.
- Access, Use and Sharing Constraints: Licensing information and any restrictions on data use.

To assure usability and accessibility, the information contents of the ICT facility concerning Raw and Processed Data are arranged as follows:

DATA DESCRIPTION

Kind of data - Common name by which the type of data is identified (ex. Raw/processed data from ground penetrating radar technology, data from seismic, magnetic, gravitational or ground penetrating radar surveys performed on ground or by using systems mounted on drone / aircraft).

Short description - Brief description of this type of data and its semantic

Data format - File format / database type used for encoding and storing.

File Nomenclature - Example of the data file nomenclature.

ACQUISITION INFO

Type of acquisition – Adopted technology and its specifics

Acquisition description – Measurement parameters

Expected size of each file - Average space occupied on the hard disk by each stored file

Acquisition continuity - Always running / Only in a period (ex. in case of temporary installation of sensors or measurement campaign)

Data release frequency - How often a file is released (es. daily, hourly, every 15minute ..)

GEO-LOCATION INFO

Data carries information on geolocation: Yes/No

Yes: each sample has its own geolocation information, which is stored inside the data itself

No: information about geolocation is not stored inside the data itself, but it is available in a separate file or not available at all.

Geolocation Notes – detailed description regarding how the data is linked to a geo-location (especially if previous answer was NO). This note reports on the link between data and geolocation information

Type of geometry - Type of geometry bound to the data, i.e. Point/Multipoint, Line/Multiline, Polygon/Multipolygon, Raster coverage. This is given both if the geolocation information is transported within the data, and if it belongs to an external file.

Map representation - 2D; 3D; image/chart/table at specific point/line/polygon

USAGE INFO

Use - Short description of the data usage

Software - software typically used for data processing (reserved to scientists) and result visualization

Discovery, sharing and interoperability info - Minimum set of descriptive elements that are useful as metadata (shortest list of metadata needed to identify the data) including information about how the data should be shared (Partners only on VPN; Web Access through Search/Download Catalog application, WebGIS access through webGIS services and webGis application).

Furthermore, the information contents of the ICT facility concerning the products are arranged as follows:

PRODUCT DESCRIPTION

Kind of product - Common name by which the type of product is identified (ex. Image, graph, map, etc).

Short description - Brief description of this type of product and its semantic

Product format - File format / database type used for encoding and storing.

File Nomenclature - Example of the product file nomenclature.

PRODUCTION INFO

Type of elaboration – included Adopted software and its specifics

Expected size of each file - Average space occupied on the hard disk by each stored file

Elaboration continuity - Always running / Only in a period (ex. in case of temporary installation of sensors or measurement campaign)

Product release frequency - How often a file is released (es. daily, hourly, every 15minute)

GEO-LOCATION INFO

Data carries information on geolocation: Yes/No

Yes: each sample has its own geolocation information, which is stored inside the data itself

No: information about geolocation is not stored inside the data itself, but it is available in a separate file or not available at all.

Geolocation Notes – detailed description regarding how the product is linked to a geo-location (especially if previous answer was NO). This note reports on the link between data and geolocation information

Type of geometry - Type of geometry bound to the product, i.e. Point/Multipoint, Line/Multiline, Polygon/Multipolygon, Raster coverage. This is given both if the geolocation information is transported within the data, and if it belongs to an external file.

Map representation - 2D; 3D; image/chart/table at specific point/line/polygon

TIME DIMENSION INFO

The product also carries information on time dimension: Yes/No

Yes: the timestamp is stored inside the product

No: the timestamp is stored aside the product (ex. In filename with a specific encoding)

Time Dimension Notes - detailed description regarding how the product is linked to a time information (especially if previous answer was NO).

USAGE INFO

Use - Short description of the product usage

Software - software typically used for product visualization

Discovery, sharing and interoperability info - Minimum set of descriptive elements that are useful as metadata (shortest list of metadata needed to identify the product) including information about how the product should be shared (Partners only on VPN; Web Access through Search/Download Catalog application, WebGIS access through webGIS services and webGis application).

The storage structure follows a hierarchical file system:

1. Root Directory: General folder structure containing major categories (e.g., Raw_Data, Processed_Data, Metadata).
2. Subfolders: Further divide each category into subfolders based on source, project, or date.
3. Database Management: Use a relational database to store structured data, while a NoSQL database is employed to unstructured or semi-structured data, allowing for efficient querying and retrieval.

Moreover, the storage structure implements tables for different data types (e.g., a table for raw geophysical data, another for processed results) with defined fields for key attributes (e.g., survey date, location, data source) and links tables through foreign keys to establish relationships between different datasets (e.g., linking seismic data to geological data).

Being metadata critical for understanding, locating, and using geophysical data effectively, the ICT facility is designed in such a way to implement a robust metadata framework accounting for adherence to established metadata standards. ISO 19139 has been identified as the minimum metadata profile based on ISO 19115, which is the standard for geographic information metadata and provides guidelines for describing geographic datasets, and the Federal Geographic Data Committee (FGDC), which is a standard for geospatial metadata, ensuring consistent documentation. Like data, also the metadata follows a hierarchical structure to facilitate easier navigation and retrieval:

1. Dataset Level: Metadata that describes the dataset as a whole.
2. Granule Level: Metadata for individual data files or subsets within a dataset.
3. Field Level: Detailed descriptions of specific fields within the data files, including units of measurement and definitions.

3. WEBGIS PLATFORM

The webGIS platform is the core component of the ICT facility because it is the tool enabling users to visualize, analyze, and manage the data. It is built to be compliant with the OGC (Open Geospatial Consortium) standards, in particular WMS (web Map services) e WFS (Web Feature Services) open-access by third party applications.

Features, components, and functionalities of the webGIS platform are summarized below.

The webGIS platform supports a multi-user environment with various profiles, each tailored to specific user needs. The identified key roles are:

1. Administrators: manage the overall system, including user access, data integrity, and platform configurations.
2. Researchers: Access and analyze datasets, upload new raw and processed data, manage metadata and table of contents for the end-users.
3. End-users (i.e., practitioners and policymakers): Access summary data and visualizations to inform decision-making processes.

The access control mechanisms ensure that users can only view or modify data relevant to their roles, protecting sensitive information.

Furthermore, the webGIS platform is designed to provide an intuitive access to the information contents. Therefore, it accounts for advanced search functionalities, such as: search by keywords (i.e., locate datasets using keywords in titles, abstracts, or specific metadata fields); filter by categories (i.e., narrow down results based on data type, date, or geographic region); faceted search options (i.e., users can exploit multiple filtering options to refine their search based on metadata attributes).

In addition, the platform offers various visualization tools to enhance the exploitation of the information contents. In this regard, the platform foresees interactive 2D and 3D maps to visualize raw / processed data, which support zooming, panning, and layer management.

4. WEB SERVICES FOR DATA ACCESS

Web services allow different applications to communicate over the web in order to facilitate the access to datasets and their associated metadata, to upload new contents and to manipulate the information, f.i. by performing operations like filtering, querying, and analyzing datasets. In this frame, the two usually adopted protocols REST (Representational State Transfer) and SOAP (Simple Object Access Protocol) have been considered. REST is widely exploited due to its simplicity and scalability; conversely SOAP is more structured and secure. By leveraging both RESTful and SOAP protocols, the platform will accommodate a wide range of user needs, ensuring that geophysical datasets are easily retrievable, manageable, and sharable. This capability enhances collaboration and data-driven decision-making across various fields, ultimately advancing research and application in geophysics.

Therefore, the ICT facility is expected to have a positive impact on the collaboration among researchers, practitioners and end-users fostering interdisciplinary projects, it will assure rapid access to existing data that is useful for improving knowledge about the territory, it will contribute to innovation opportunities.