MISSIONE 4
ISTRUZIONE RICERCA

Deliverable 7.8 (B12, Activity 7.8)
Implementation of new geophysical data management system with dedicated storage and computing facilities

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

• Deliverable number:	• D7.8
• Work package:	• WP7 – 7.8
• Intermediate Objective:	• IO7.5
• Deliverable type:	• X Document, report
	• X Websites, patent filings, videos, etc.
	• ☐ Other: please specify .....
• Dissemination level:	• X Public
•	• ☐ Restricted
• Estimated delivery (bimester):	• B12
• Actual delivery date:	• B12
• Author(s) (Partner-OU):	• Paolo Diviacco, Alberto Viola
• 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

Deliverable 7.8 reports on the SNAP web based data system able to handle geophysical data and in particular Exploration Seismics in the context of the Itineris Projects.
Exploration Seismics is one of the most important geophysical methods that could provide insights of the Earth crust up to depths of several Kilometers. This approach has been used widely in many areas of the globe accumulating large datasets that allow to improve the knowledge of the Earth dynamics. Exploration seismics have specific characteristics when compared to other type of geophysical data. Data need to be georeferenced but differently from other data types this type of data is expressed through seismic profiles. These consist of vertical slices of the Earth interior that normally is expressed not in depth but in time. This is due to the fact that what is recorded at the surface is the timing of reflections in the underground which depends on the velocity of the materials where the elastic perturbation has travelled through. If this resembles the data products of seismology it is instead rather different because the representation is not an isolated signal as a function of time but rather an alignment of vertical responses of that specific location.
This poses specific problems from the point of view of web-based visualization that the SNAP infrastructure solves. In fact, no web GIS can address such problems while a specific solution is needed. At the same time if some commercial solutions are available, they are available only to large companies and cannot be customized to the need of research institutions.
The main objective of this deliverable is to provide a general overview of the structure and contents of the SNAP platform.

2. SEISMIC DATA NETWORK ACCESS POINT (SNAP)

Seismic data Network Access Point (SNAP) is the data web portal and framework that allow to access all geophysical data acquired by Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS. SNAP hosts approximately 91,073 km of seismic lines and 351,000 square kilometres of Multibeam data, mostly in the Mediterranean and Black sea and is complemented by the Antarctic Seismic Data Libray (SDLS) that under the auspices of the SCAR committee gathers all the seismic data (approx 300,000 km) acquired by all the international research centres in Antarctica.

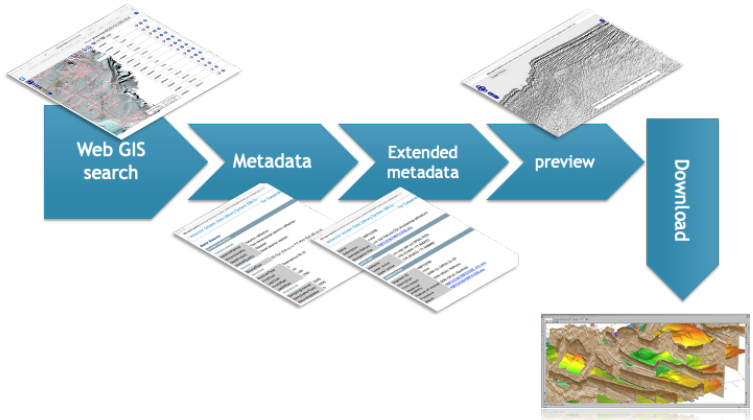

Figure 1Workflow of data access to Geophysical data in SNAP

2.1 Data Standards

Data standards in exploration seismic have always been conditioned by commercial traditions and practices. Considering that seismic data correspond to large arrays of values, they are not stored within a database but rather as files. The most commonly used standard for storing and exchanging seismic data is the Society of Exploration Geophysics (SEG) SEG-Y format. Through time the SEG-Y standard has remained quite rigid, even if new technologies emerged. The pressure to accommodate these new requirements forced many dialects and interpretations of the same data structure, that came into conflict when different teams and communities interacted. The most problematic issue relates to the use of the SEG-Y trace headers. These are structured blocks of 240 bytes positioned within the seismic data file, at the begin- ning of each seismic data trace, and holding information such as trace length, sampling interval, etc., but no provision exists for example for storing floating point values. Within SNAP data files are uploaded with the minimal set of needed information while leaving the less standardized information to external metadata that we have semantically defined in several previous works
An example of the use of an ancillary file is positioning data, for which we adopted an external ASCII file in the UKOOA P1/90 for- mat that was defined by the UK Offshore Operators Association in the early 90s for use in the oil industry. This format has been adopted in SNAP as it allows a minimal set of required information to be loaded.

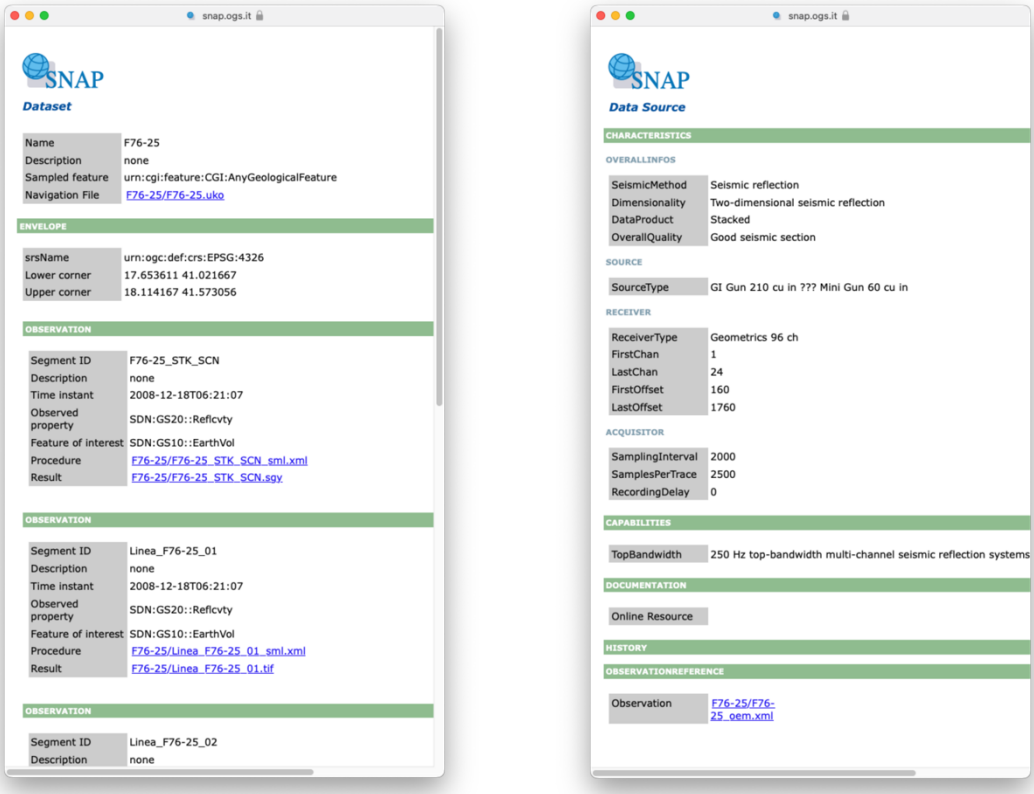

Figure 2 OGC compliant O&M and SensorML Metadata used within SNAP

2.2 Metadata

Within SNAP, metadata is used both for data usage, meaning processing, integration or interpretation, and for data discovery. The same parameter, such as, the seismic data sampling frequency can be needed on one side to set a frequency filter and on the other to understand whether

a dataset can be used for engineering or for scientific research. This information needs to reside in multiple locations: it must be contained in the data file that can be downloaded from SNAP for further processing; it must reside in the database on the top of which the web interface is run, and, at the same time it must be available to other initiatives that need to know what kind of data SNAP can offer. Not all of these applications need the same set of information. Processing needs data related information, discovery needs contextual information. The SNAP discovery facility relies on and queries a database that contains all the relevant metadata and links to the data. Most of the metadata have been extracted automatically from seismic data during upload in the system, while several parameters still need manual intervention. Of course the internal intricacies of the system are hidden from the end-user who simply searches and finds what he/she is looking for from the web interface.

2.3 Data Rescued

Data rescue requires extraction of data from a storage system that no longer guarantees their preservation, to another system that is, or at least currently thought to be, more reliable and accessible. OGS archives have been built up over an extended period of time, mostly as an extended collection of tapes and paper printouts. 21-track tapes were initially used in the 1970s, then standard digital 9 track tapes, and more recently, the 3480/90 and exabyte tape cartridges became the norm. The institute had no funding to maintain tapes in controlled storage conditions, so all media underwent a progressive deterioration, the effects of which were particularly severe on the old 21 and 9 track tapes. Paper prints date back to the seventies and were produced using photographic or electrostatic paper. Their archival stability depends upon processing, display and storage conditions. In the case of the OGS archive, although not optimal, these were not dramatically detrimental. Photographic paper prints did not undergo severe deterioration while electrostatic paper prints were more of a concern, most of them had no simultaneous digital copy. Paper sections and maps were first digitally scanned using large dimension (A0) scanners to obtain high resolution (300 dpi, a software factory preset that can- not be changed) raster files. Later, seismic profiles were converted to actual (SEG-Y) digital seismic files while maps were georeferenced and navigation or positioning digitized manually

3. THE NEW WEB PORTAL

Within SNAP, seismic data can be searched geographically and textually. Once data are found it is possible to read their metadata and download positioning in a UKOOA file. Data can be visualized either as a static image file or as actual Seg-Y data, where basic processing on the data is possible. This workflow is represented in Figure 1

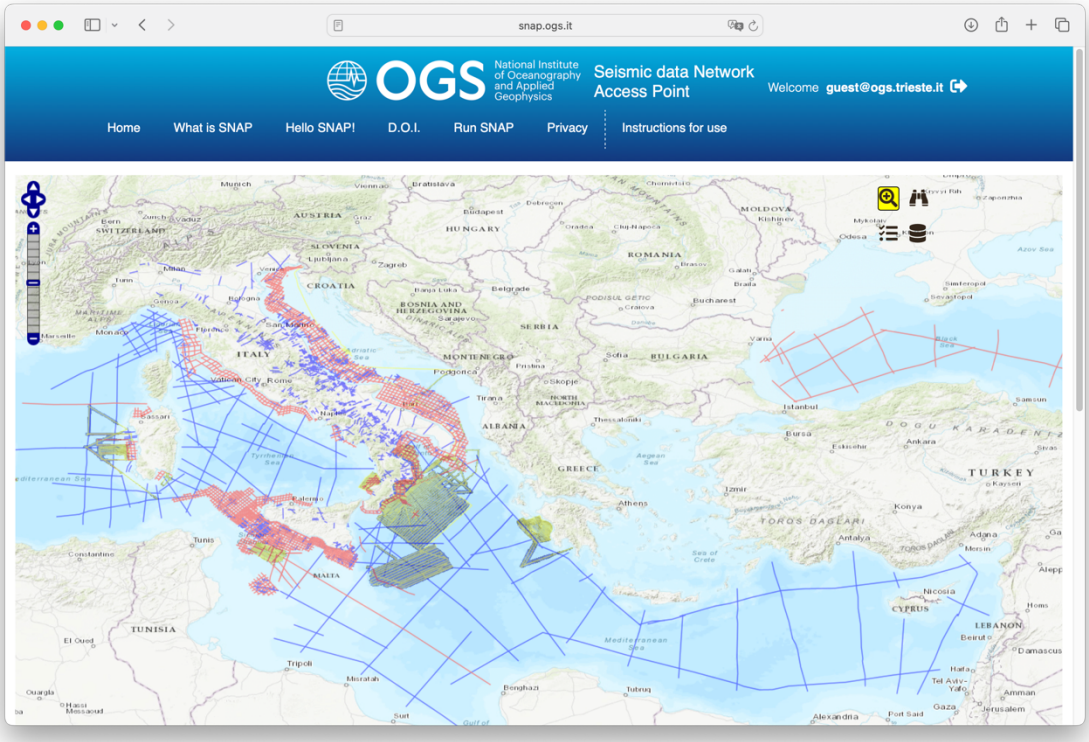

Figure 3 The web-GIS interface of the SNAP portal that allows to identify the seismic data users can be interested in

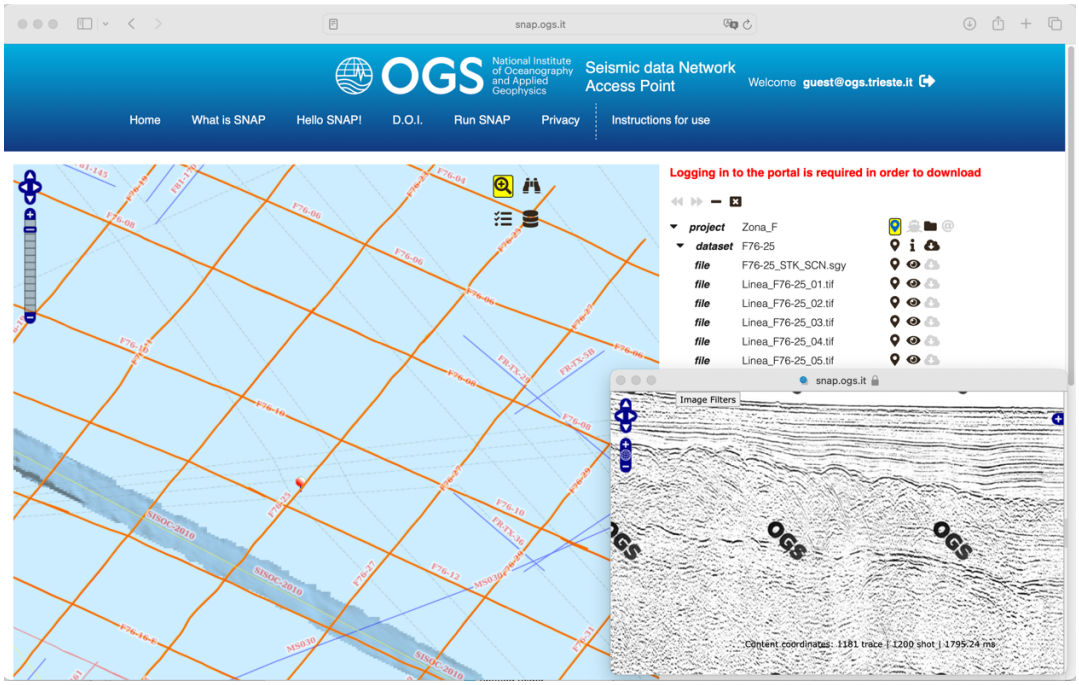

Figure 4 Once selected Data can be accessed interactively and previewed on line

3.1 The structure

The site was created using a combination of standard web technologies and custom solutions:

- Hosting: The site is currently hosted on the internal server infrastructure of the National Institute of Oceanography and Experimental Geophysics-OGS.
    - The site runs on Virtual Machine (KVM version 8.2.2) on multiple and mirrored Dell R840 (Intel Xeon Gold 6252N CPU @ 2.30GHz, RAM 256 GB, disco 1787.5 GB [2 SSD-SAS RAID-0]) servers, with Ubuntu 24.04.1 LTS;
    - Storage is granted by Dell EMC ME5024 Storage Array SSD
    - Power supply is granted by Mulitple levels UPS systems ;
    - The system embeds also metadata and harvesting facility based on geonetwork software
    - The system is mirrored as portal and as data using a redundant array of storage systems powered by a SNAP developed back up script
- Technologies used:
    - Java + Apache Tomcat: For server-side logic and dynamic content generation;
    - HTML + JSON: To structure the content of web pages and data interchange;
    - Bootstrap Italia: A front-end framework specific to Italian public administrations, which provides predefined components and styles that are compliant with the design guidelines for PA websites, ensuring visual consistency and accessibility.

The site is integrated into the OGS site and has been created within the frame of the coordinated image of OGS, with colours, fonts, and style in agreement with the OGS principal site and the other hosted ones. In Figure 1, the home page.

The portal is organized in 7 sections:

- **Home**: a rapid overview of SNAP's mission and functionalities with a menu to access its functionalities;
- **What is SNAP**: An introduction to the framework with references to pubblications that describe its philosophy;
- **Hello SNAP**: a dynamic map locating users of SNAP;
- **DOI**: lists all the landing pages related to the DOI assigned to data stored within SNAP
- **Run SNAP**: The launcher of the web app that allows to select and access the data;
- **Privacy**: GDPR notice and description of how data are used within SNAP
- **Introduction for use** : a quick tutorial on how to use the web app to access the actual data, select them, preview and download them

3.2 SNAP interoperability through Geonetwork

SNAP embeds GeoNetwork in order to offer interoperability services such as harvesting to other data sharing initiatives such as, for example, INSPIRE. GeoNetwork is a free and open-source software, a catalog application to manage spatially referenced resources. The software is widely used in numerous Spatial Data Infrastructure initiatives this is because interoperability and searching through multiple catalogs is one of the strengths of this application.
GeoNetwork supports both DCAT and ISO19115/119/110 metadata standard formats.
The online editing of metadata is based on a template system and directories of information. A dataset is therefore described on the basis of a template, a metadata scheme, whose fields are filled in through the use of specific terms derived from the thesauri loaded into the catalog. Moreover, once a schema has been compiled, through GeoNetwork it is possible to verify that the metadata created is INSPIRE compliant.
Given the potential of GeoNetwork and given the importance of managing the data according to The FAIR Principles (Findability, Accessibility, Interoperability and Resuability) SNAP uses GeoNetwork to generate metadata to complement the metadata already present within SNAP. In this way, at the level of the GeoNetwork catalog, the metadata is more generic in nature, and, through

the use of the DOI, it is possible to get to the domain-specific metadata present in SNAP, the O&M and SML.

SNAP implementation of GeoNetwork uses the following list of thesauri:

Conditions Applying To Access and Use
http://inspire.ec.europa.eu/metadata-codelist/ConditionsApplyingToAccessAndUse
Continents, countries, sea regions of the world.
http://geonetwork-opensource.org/thesaurus/naturalearth-and-seavox
GCMD – Locations
https://gcmdservices.gsfc.nasa.gov/kms/concepts/concept_scheme/locations
GCMD – Providers
https://gcmdservices.gsfc.nasa.gov/kms/concepts/concept_scheme/providers
GCMD - Science Keywords
https://gcmdservices.gsfc.nasa.gov/kms/concepts/concept_scheme/sciencekeywords
GEMET - INSPIRE themes, version 1.0
http://inspire.ec.europa.eu/theme
INSPIRE glossary
http://inspire.ec.europa.eu/glossary
Limitations on public access
http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess
Topic categories in accordance with EN ISO 19115
http://inspire.ec.europa.eu/metadata-codelist/TopicCategory
W3C DCAT Themes for the SeaDataNet EDMED Catalogue
http://vocab.nerc.ac.uk/scheme/EDMED_DCAT_THEMES

3.3 Future work

A Joomla-based version of the site is under development with the following objectives:

- Simplify page maintenance;
- Allow non-technical users to manage content through a user-friendly interface;
- Leverage the Joomla ecosystem of templates, plugins, components and modules to extend the site's functionality;
- Potentially improve SEO performance and content management.

The migration to Joomla will allow for more efficient site management in the long term, reducing dependence on specific technical skills for routine content updates.

Moreover, the site will be upgraded with links to the new portals under development and constantly updated with the latest projects and news.

4. REFERENCES

Diviacco P, Busato A. The Geo-Seas Seismic data viewer: a Tool to facilitate the control of data access. Bollettino di Geofisica Teorica ed Applicata 2013;54/2.

Diviacco P, Lowry R, Schaap D. Marine seismic metadata for an integrated European scale data infrastructure: the FP7 Geo-Seas project,. Bollettino di Geofisica Teorica ed Applicata 2013.