DigLogs

Functional and Technical requirements

Deliverable D5.3.3

5.3.3. Spatial Data Management System

Responsible partner: CFLI (PP1) **Involved partners: All**			
Version	**Status**	**Date**	**Author**
1.0	Final	10.06.2021	CFLI
Notes:			

Table of Contents

1 Introduction

1.1 Pilot project purpose

The pilot action "Spatial Data Management System" is to be implemented in the context of the North Adriatic Sea Port Authority and it regards the adoption of a **centralized and interoperable spatial data repository** aimed at giving a robust structure to the information and data used within the internal processes and to provide services to external operators and institutions.

Belonging to the innovation named "Maritime Big Data / Data management", it aims at obtaining the best results from integrating different data sources in terms of added value in knowledge and management capability. The specific objective of the pilot action is to enable an integrated management and utilization of standard data, real time data and georeferenced (spatial) data both to support decision making processes and improve Port Authority services overall quality.

The new system implementation is mainly aimed at making a transition from a current situation in which data is ineffectively managed and used to an improved condition in which more different data can be integrated and dynamically accessed by several users according to different policies and objectives without replication and corruption. Indeed, in the current situation of Venice Port Authority, the decision-making support based on the use of spatial data is provided copying and processing several times the main datasets due to the utilization of different storage systems and processing techniques within the same organizational context. This makes difficult both to keep dataset up-to-date and share it in an effective way, and it significantly reduces processes performance.

For the innovation deployment, a Spatial Data Infrastructure (SDI) will be implemented mainly to store and easily share all the in-use geographical datasets. It will perform both the storage and dynamic processing functions, making different users able to directly access data and processing results and visualization according to a special policy management protocol. This Spatial Data Management System will allow to store the processed data and maps either as new datasets or as algorithms that process data in real time, without forcing operators to change the already known working tools.

The SDI implementation won't need any software development; therefore the pilot action will have a "training-empowered" approach in order to achieve both an organizational improvement

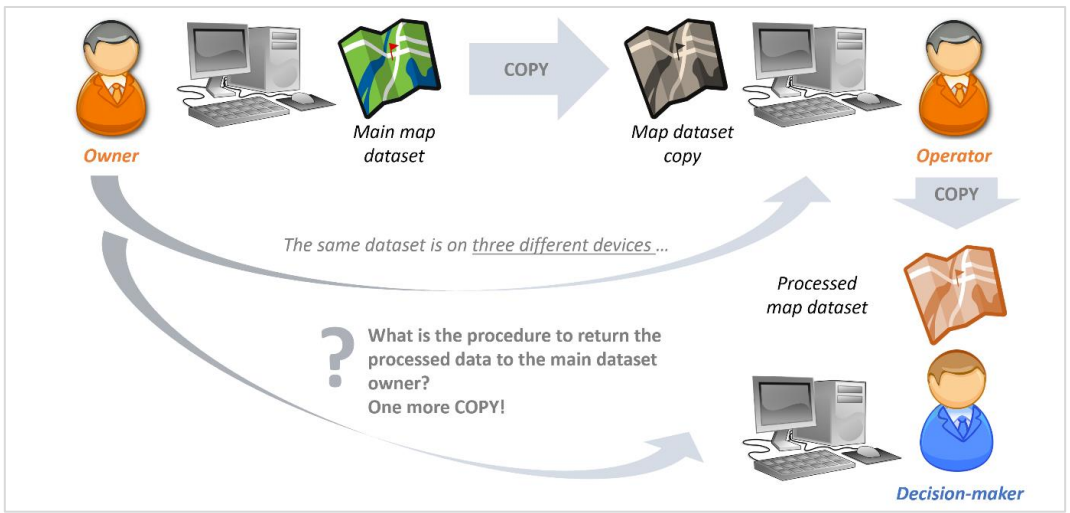
and a workforce skill improvement, fostering the awareness on how spatial data visualization and dynamic data processing can support decision-making process. For this reason, the educational and training programs are a substantial part of the pilot implementation.

1.2 Target

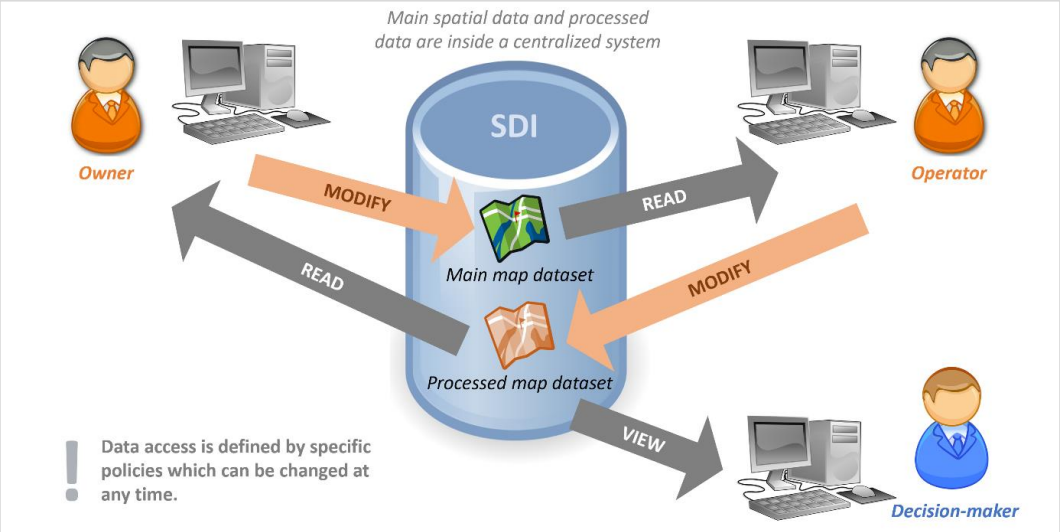
The pilot implementation tasks are aimed at implementing a new Spatial Data management Infrastructure (SDI) at the Port System Authority of the Northern Adriatic Sea Ports of Venice and Chioggia.

From an IT point of view, the new SDI include two components: a geodatabase and a map server; the first one is oriented to storage and process geographic data in a centralized way, the second one will provide access to the geographical data through OGC interoperability standard according to the international specifications.

From an operational point of view, the mid-term implementation perspective includes several activities aimed at migrating of the in-use geospatial datasets, currently managed with low-efficiency, non-standard and non-interoperable systems, within a new more effective system. Furthermore, in parallel with the technological implementation, special activities are aimed at redesigning the data-driven decision processes and at raising the skill level of the organizations' employees.



Port Authority data management process current situation



Port Authority data management improved situation

In the context of the infrastructure deployment, also a software applications review is needed in order to identify the need for reconfiguration, updating, integration or replacement of some of the in-use tools and systems.

1.3 Scope

The main pilot function is to provide the Venice Port Authority with a new IT integrated system for the management and utilization of standard data, real time data and georeferenced (spatial) data both to support decision making processes and improve Port Authority services overall quality.

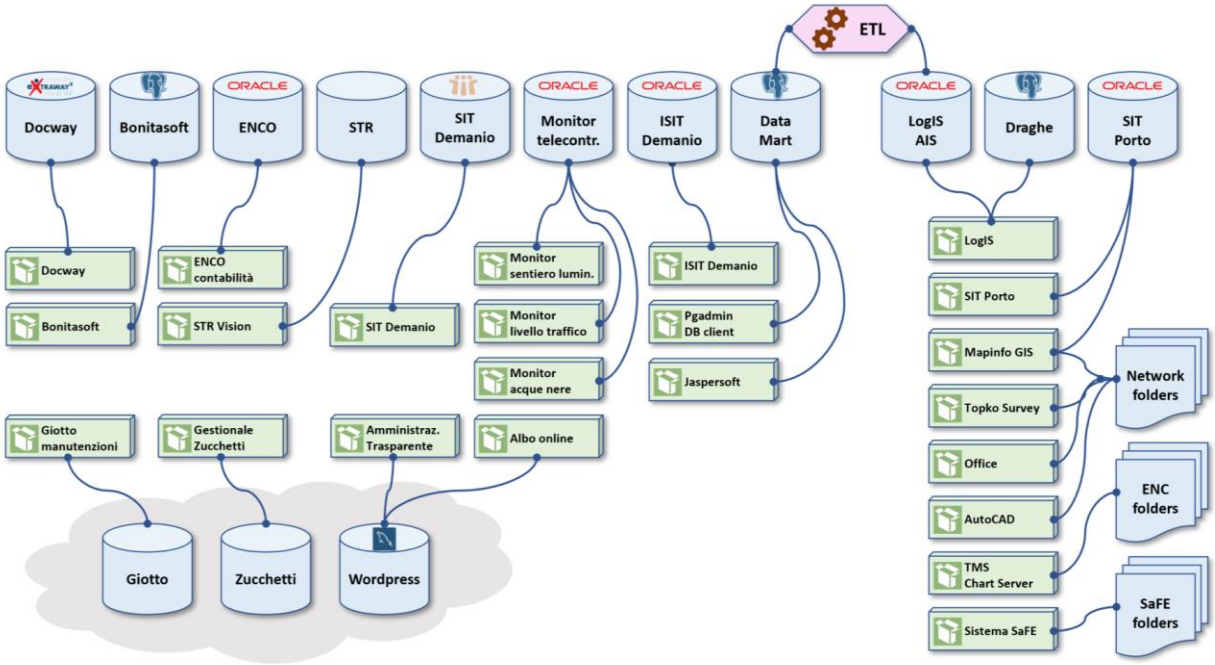
The pilot scope is to make a transition from a current situation in which data is ineffectively managed and used to an improved condition in which more different data can be integrated and dynamically accessed by several users according to different policies and objectives without replication and corruption.

With the pilot action, a **Geodatabase** will be installed and configured as the main engine of the **Spatial Data Infrastructure** and **operators will be trained** to store and process data within the

new system. The Geodatabase engine will allow all Port Authority operators to visualize and process structured and optimized spatial datasets, and create maps using data produced and managed by all internal department areas. Indeed, the operators will be able to connect their workstations to the SDI using their account and use all data to produce desired map outputs. The Spatial Data Management System will allow to store the processed data and maps either as new datasets or as algorithms that process data in real time, without forcing operators to change the already known working tools.

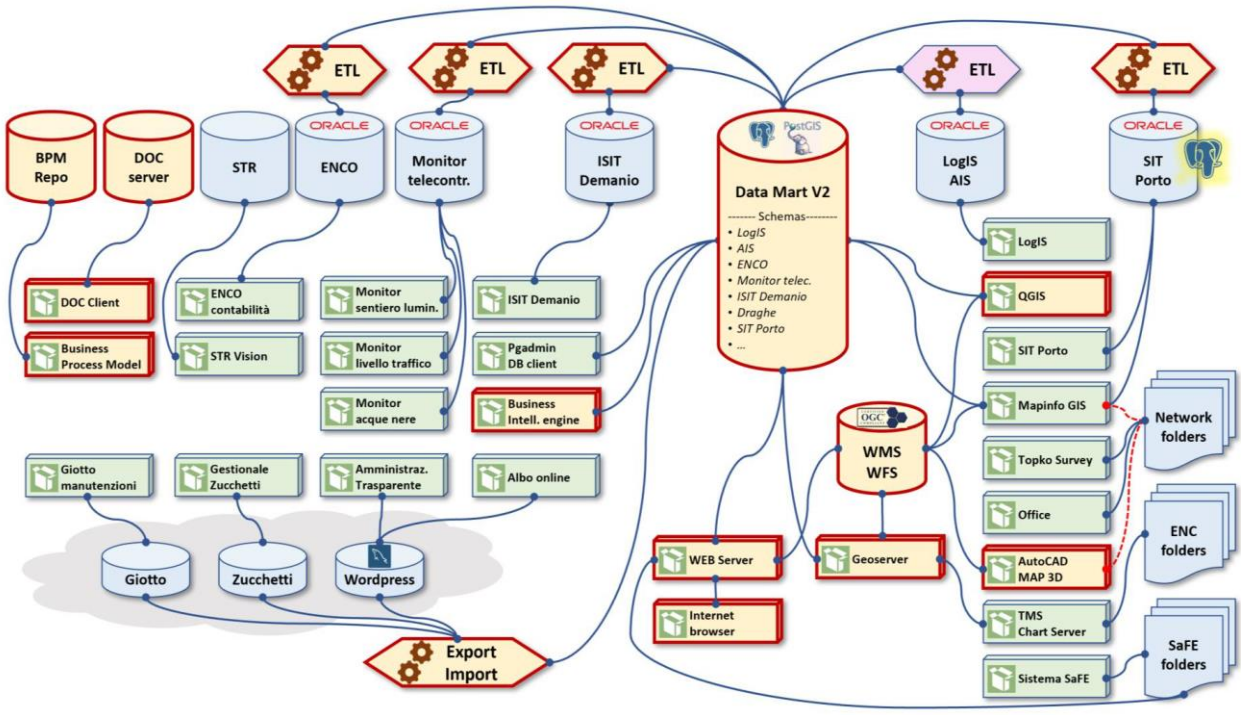
The SDI implementation won't need any software development, therefore the pilot action will have a "training-empowered" approach in order to achieve both an organizational improvement and a workforce skills improvement, fostering the awareness on how spatial data visualization and dynamic data processing can support decision-making process.

The data repository and software tools **as-is implementation scenario** is the one shown in the following diagram:



This first diagram shows the Venice Port Authority current situation in terms of data management systems (cylinders), applications (boxes) and computer folders (folder packets on the right). The main issue of this scenario is related to the high fragmentation of the whole system.

A global reorganization of the entire system needs a special implementation plan that can be drafted taking into account some possible interventions as shown in the following diagram:



In the possible long-term reorganized scenario, some news items will be added (red-edged/yellow items), but the main proposal is related to the implementation of a new data management system called “Data Mart V2” which will include a geodatabase engine able to manage in a unified way most of the datasets.

Other proposed/possible components of the improved system will be new Database Management Systems needed due to the replacement of the BPM and Document Management System, a new interoperable mapping server and related web components, as well as some ETL

engines needed to synchronize the main geodatabase with the ones in use by LogIS Port Community System and other dedicated management software.

As always happens, the architecture of such a system is quite complex, so a step-by-step approach is recommended in order to progressively implement the system avoiding the disruption of existing processes and services.

2 Mid-term implementation plan

A mid-term implementation plan will be required to outline the development future scenario after the pilot action completion, and it will be designed based on a previous "as-is" analysis of the available data assets and hardware/software tool drafted for the Interreg project SUPAIR during 2018.

Although the mid-term implementation actions overcome the DigLogs Pilot Action execution, the implementation plan details and estimates can be considered the part of the pilot itself that ensure the proper resource allocation for the future development of the system.

2.1 Objectives

The mid-term implementation plan will include the following objectives:

1. Implementation of the **datasets managed by the "Pianificazione Strategica e Sviluppo" department** inside the new Spatial Data Management system
2. Implementation of the **datasets managed by the "Direzione Tecnica" department** inside the new Spatial Data Management system
3. Implementation of the **datasets managed by the "Affari Istituzionali e Demanio" department** inside the new Spatial Data Management system
4. Implementation of the **datasets managed by the "Programmazione e Finanza" department** inside the new Spatial Data Management system
5. Adoption of a platform to manage **metadata** and interoperable services
6. Analysis, maintenance and updating of **existing software tools**

2.2 Actions on datasets

Time and costs estimate regarding actions on the datasets will include the tasks listed below; each task may be assigned to one or more specialists or companies with proper technical and scientific skills:

- Data **collection** (extraction from the in-use storage system)
- For multi-part complex datasets, **extraction of the basic component datasets** as well as identification of internal relationships, needed to perform the next steps (for time/costs estimate, a 500% multiplier factor is used)
- **Analysis** of the structure and contents **quality assessment**
- Identification of **possible existing keys** to be used to join related datasets, integrity check and optimization if needed
- **Merge** of different dataset parts if needed
- Definition of the proper **use mode and protocol** (accessing, visualization, processing) according to the dependent processes, visualization and processing procedures definition and test
- Definition and test of a proper and effective **maintenance protocol**
- For all the table datasets, identification of possible **layers to be used as geocoding** base and test of the needed procedures
- **Loading** of the dataset into the new Spatial Data Management System and access test
- **Reporting** of the needed information to access, visualize, relate, process and maintain the dataset
- **Education and training**

2.3 Actions related to processes

Time and costs estimate regarding actions for processes redesign will include the tasks listed below; each task can be assigned to one or more specialists or companies with proper technical and scientific skills:

- On the basis of the **as-is analysis details**, identification of both datasets managed by each process (that is datasets in which the process performs insert, modify and delete operations) and read-only access datasets

- As regards the managed datasets, **identification of the in-use procedures** for information collection, processing, possible replication and other actors involved in the data management
- For all datasets, **verification of the in-use storage systems**, effectiveness and efficiency of the **access and processing tools and procedures**, and the used/usable formats
- Analysis of the needs of special **processing procedures** (data derivation, integrated processing, data correlation between datasets …)
- Analysis of the need of special **user functions or interfaces** to be implemented by developing new software procedures, new stand-alone applications or applications to be integrated with existing ones
- Definition, implementation and test of special **protocols** to perform required processing, storage and updating operations
- **Education and training**

3 Functional and system requirements

3.1 Spatial data redundance and fragmentation reduction

In the current scenario, it is quite impossible to unitarily manage almost all available datasets due to fragmentation, duplication, redundance and lack of reference about ownership and responsibility for the data content. This issue will probably worsen as new survey campaigns are underway, such as the new aerial photogrammetric survey or the Chioggia port area underground services survey.

For these reasons, the first functional requirement is related to the reduction of the fragmentation of the ICT system and the redundance of data sources. For this action, it will be required an adequate list of data sources with which to perform the optimization and migration operations and implement the base core of the new Spatial Data Infrastructure.

For each identified data source will be performed the following operations:

- Acquisition and quality assessment
- Update and ownership verification

- Merging, cleaning and structure optimization
- Conversion if needed
- Migration

3.2 Grant policy definition

The grant policy will be defined on the basis of the Port Authority internal department areas subdivision, ensuring read/write permission on datasets to the user belonging to the area that manage them and read-only permission on all datasets to all users.

For the raster segment of the SDI, no special permissions will not be needed as data access will be performed through standard interoperable protocols in read-only mode.

Grants will be applied to user groups instead of single user, in order to ensure flexibility in the assignment or re-assignments of users to different areas.

3.3 Direct and flexible use and re-use of spatial data

The SDI will allow all Port Authority operators to directly access in real-time to native data source which could be updated by owners in a concurrent non-conflicting mode. This will be done according to the predefined grants policy.

3.4 Customized support and training

The SDI will be available for direct use to all Port Authority employees based on an accounting system and a predefined grants policy. This will be possible using different type of client application, for both spatial and table data sources, through the application of the current interoperability standard protocols.

The broader utilization of the infrastructure will be ensured thanks to a special and customized support and training programme. To achieve this goal and enhance user performance in using the SDI for working tasks, it will be required an identified and motivated group of employees to participate in the programmed training activities.

3.5 New aerial photogrammetry survey

In the coming months, a new aerial photogrammetric survey will be carried out providing a new vector base map and a high resolution orthophoto for the whole port area. The new digital base maps will be rich, detailed, and multi-layer, and they must be optimized for wide access within all the Port Authority departments. The data model definition, the optimization and the visualization fine-tuning currently seems to be very complex to be carried-out by port authority employees.

The SDI will store the new survey dataset in order to ease the access for all Port Authority operators using the same procedures as those for all other data sources.

3.6 Big datasets management

Some available and future datasets will be difficult to be managed due to their size; those are for example the bathymetry surveys datasets in ASCII point cloud vector format and the high resolution orthophoto that will be provided in the coming months. For this kind of datasets, it is usually impossible to uniformly visualize the total covered area or process more than one dataset at a time.

Special conversion procedures and a raster data segment will be needed in order to provide effective access to huge datasets through optimised web protocols.

4 Hardware/Software requirements

4.1 Hardware requirements

The SDI will need a physical or virtual server with the following minimum requirements:

1. Operative System: Linux or Windows
2. CPU: 4-8 cores at 2.x-3.x Ghz (this affects waiting/response times during visualization and geoprocessing)
3. RAM memory: 16-32GB (this affects the capability to manage huge datasets, especially raster data such as orthophotos, with a mapping server, or the capability to manage

concurrently requests coming from different users). Usually, for a Windows server more RAM is recommended.

4. Storage: 2TB upgradable if needed (affects the capability to store more datasets, bigger datasets, backup, replication and historicization of data archives). Additional storage is to be defined in order to store the orthophoto coming from the new aerial survey.

4.2 Software requirements

1. Vector data segment implementation (geoDBMS)
    a. PostgreSQL version 10 or later with PostGIS spatial extension
2. Raster data segment implementation (OGC compliant mapping server)
    a. Geoserver latest version (currently 2.17)

5 Spatial datasets for pilot implementation

As result of the pilot action pre-assessment analysis, the spatial dataset ready to be integrated into the new Spatial Data Management System are the following:

1. Aerial photogrammetry digital map
2. Orthophoto
3. Coastline boundary
4. Weather station data
5. Buoys / Dolphins
6. Mo.S.E. system structures
7. Navigation aid systems
8. Railway
9. Streets
10. Venice Port boundary
11. Trees and green areas
12. Waste bins location
13. Seabed bathymetry
14. Lighted pathway

15. Public property areas
16. Canals and basins
17. Underground infrastructures

NOTE: implementation of dataset 1, 2 and 3 is conditioned by the availability in time of the aerial survey results.

6 Training and educational program requirements

6.1 Content reference framework

6.1.1 Training module types

The following is the content reference framework to be used for the definition of the detailed training and educational program.

- ❑ Seminars

  Mainly theoretical contribution aimed at deepening political-cultural, informative, awareness-raising aspects or, in more specific cases, understanding of the potential and limits of some technological or methodological approaches. Contents are divided in three main categories: ICT reference scenario, territorial issues, technological in-depth.
  Optimal duration; 1 hour

- ❑ Workshop/training-on-the-job

  Training modules carried out with the use of computers and software aimed at providing specific skills in the use of data processing tools and methods. The aim is not to train a specialist but to make aware of advantages / disadvantages / criticalities of the use of certain tools and methods, as well as introducing the participants to the use of new tools and allow them to self-train according to the future needs.
  Contents are divided in six topics: Spatial data management, Data Base, GIS data processing, Remote Sensing, Web / geospatial-Web, Info design and spatial data visualization.

Optimal duration: variable 8-20 hours for 1 to 5 days, depending on the needs and content.

- ❑ Project work
  Mainly a strategic design type educational activity, with possible small technical insights. In the most effective formulation, it provides joint and synergistic activity between decision-makers and executives by simulating the work of a multidisciplinary team. Optimal duration: variable from 4 to 8 hours within one day.

6.1.2 Seminars content reference list

The following reference lists are intended as basic ranges of topics useful for the definition of a custom educational path based on the specific needs of the participants.

ICT scenario:

- *Smart Cities & Communities / Sustainable Cities*
- *Geography and knowledge about territory: ICT based approaches*
- *Data interoperability, Open Source and Open Data approaches*
- *Big Data and Data Analytics*
- *Mobile APPs and Mobile Web*
- *IOT – Internet of Things*
- *Internet, knowledge sharing and participatory processes*
- *Institutional data sources and spatial data: the regulatory scenario*

City, environmental and territorial issues

- *Hydraulic and Hydrogeological risk management*
- *Biodiversity protection*
- *Energy and sustainability*
- *Logistics and transportation*
- *Tourism, landscape, and cultural heritage*
- *Sustainable land use*
- *Territorial planning and regulations*
- *Technological network and infrastructures maintenance*

- *Urban life quality, comfort, and health*

ICT technologies

- *Geographic Information System and Spatial Data Infrastructures*
- *Geodatabase Management System*
- *Geo-Web solutions*
- *Big Data and Data visualization*
- *Geostatistics*
- *Terrestrial and marine sensing*
- *Satellite and airborne remote sensing*
- *Satellite active sensors monitoring*
- *Thermographic monitoring*
- *Mobile APP and Web application for geospatial data*
- *Wireless Sensor Networks and IOT*
- *Mobile Mapping Systems for road networks maintenance*
- *GPS, Laser scanning and UAV survey and monitoring*
- *Information Design and Data Visualization with geospatial data*

6.1.3 Workshop/training-on-the-job content reference list

- Cartography and spatial data management:
  Introduction to Geographical Information Systems, digital cartography, institutional data sources, geocoding techniques.

- Database
  Database design basis, SQL, DBMS applications, geospatial extensions.

- GIS. Geographic Information System:
  Basics of geospatial data modelling, visualization and thematic mapping techniques, data integration techniques and geoprocessing.

- Remote Sensing:
  EOS - Earth Observation Systems introduction, remote sensing data types and formats, raster data processing techniques, raster data classification.

- Web / Geospatial-Web:
  Web browsers, web services, basics of HTML, CSS, Javascript, mapping servers, geographic web services and interoperability protocols.
- Info Design and spatial data visualization:
  Information Design and User Centered Design, data visualization techniques and tools, mapping tools design, basics of Interaction Design.

6.1.4 Project work content reference list

- Air, water, and soil pollution risk
- Hydraulic and hydrogeological risk
- Mobility, Infomobility and logistics, road safety, intermodality and transportation
- Biodiversity protection
- Production and sustainable use of energy sources
- Protected areas management
- Cultural heritage enhancement and sustainable tourism
- Emergencies management
- Urban quality and services
- Waste management
- Infrastructures, green areas, and services maintenance

7 Spatial Data Infrastructure requirements

7.1 Data types

The SDI system will allow to store and manage the following types of datasets

- Geospatial layers (point layers, linestring layers, polygon layers, in both single-part and multi-part version)
- Alphanumeric tables (standalone tables, layer-related jointing tables)

- Rasters datasets (georeferenced remote sensing and aerial images, DEM, DSM and other non-optical raster datasets)

7.2 Interoperability and Open Geospatial Consortium (OGC) compliance

The SDI system will expose the following geospatial OGC compliant web services

- Web Map Service (WMS) for visualization purposes
- Web Feature Service (WFS) for vector processing purposes
- Web Coverage Service (WCS) for raster processing purposes

7.3 Database model

The SDI vector and raster segment will be organized based on the following Venice Port Authority department area subdivision:

- DPSS_STRA (Pianificazione Strategica)
- DPSS_URB (Pianificazione Urbanistica)
- DPSS_TRAFF (Studi Marittimi e dei Trasporti)
- DPSS_COM (Ricerca e Sviluppo Progetti)
- DPSS_ICT (Gestione Operativa Servizi IT)
- DTEC_CAN (Canali e Navigabilità)
- DTEC_AMB (Ambiente)
- DTEC_APR (Progetti)
- DEM_BENI (Valorizzazione Beni Demaniali)
- DEM_CONC (Concessioni Demaniali)
- DCOP_PORT (Lavoro e Coordinamento Operativo Portuale)
- DCOP_SIC (Sicurezza, Vigilanza e Controllo)
- DPF_GARE (Gare e Lavori Pubblici)
- DPF_GEST (Qualità, Anticorruzione e Trasparenza)
- DPF_MEPA (Acquisti servizi e forniture)
- DPF_RAG (Ragioneria, Contabilità e Aspetti Fiscali)

7.4 Network

Server application standard TCP-IP ports:

- PostgreSQL: 5432
- Geoserver: 80, 8080

8 Data migration requirements

Special data migration processes and techniques will be applied to ensure the following requirements:

- Define data source ownership in order to choose the related database schema for the storage. (NOTE: ownership means either user created the data or that the user gathered the data from external institutions or companies, and it ensure the needed update/reload over time)
- Spatial dataset is based on one of the three basic geometric primitives: point, line, area, in multi-part mode (except for points that could be single or multi)
- Check geometry validity and repair it if needed
- Assign the chosen Coordinate Reference System if needed or check its validity
- Reduce at minimum amount the number of layers/tables that shares the same data model
- Check redundancy and/or duplication and eliminate them
- When possible, split extensive descriptive part from geometric part joining them by an alphanumeric key
- Define a predefined visualization style and store it into the database
- Perform format conversion if it ensures best performance
- Vector datasets are stored into the geoDBMS while raster datasets are stored in a dedicated network folder in geoTIFF format; the mapping server reads geoTIFFs and

exposes WMS and/or WCS web service. When needed, multiple geoTIFFs are exposed as a single layer using ImageMosaic plugin.

9 Procedures implementation requirements

9.1 Grants policy management

The grants policy management system will allow:

a. Group roles management
b. User roles management
c. Manual and automatic (triggered) grants policies to database items

The grants policy definition, based on the department organization in Venice Port Authority will be performed according to the following 5 steps:

- Group roles definition
- Login roles definition
- Read-only privileges setup for “public” schema
- Read-only privileges setup for all department schemas to all group roles
- Full privileges setup for department schemas to related group role

9.2 Dataset management

9.2.1 Point cloud data source

Point cloud datasets must be intended as “raw” / non-optimized datasources. The bathymetry surveys datasets are provided in this format and, since they can be considered as “morphological” models, they can be effectively converted into a more performing raster DEM format (float values 24bitg+ geoTIFF).

9.2.2 Raster dataset mosaicking

Extensive raster data will be merged into a virtual mosaic (QGIS *.vrt or equivalent).

In case of the bathymetry, when creating virtual raster mosaic files, the pixel dimension will be the same of the higher resolution of the merged files.

It is NOT recommended to perform a physical merge of the raster DEMs, since it dramatically reduces performance and usability.

9.2.3 Segmented data modelling

Geospatial datasets with large number of alphanumerical attributes will be split into two or more datasets, one geospatial dataset with minimum required attributes and one or more additional table datasets joinable by using a key field.

This is called “segmented data modelling” and it is recommended also in case of one-to-many relation between map and table items or in the case when geometric data source origin is different from the table content and there is the need to maintain them separately.

9.3 Information design, visualization and delivery

9.3.1 Light web-based visualization tools

It will be possible to create different light web viewers in HTML format consuming the spatial web services provided by the SDI.

The code of the web viewer will be based on OpenLayers or similar open-source map libraries and could be explained to operators and/or documented into a guide.

9.3.2 Layer styles management

It will be possible to define a default visualization style for each spatial dataset and store it into the SDI database. The default style will automatically be applied by default once the layer is added to a GIS project.

9.3.3 Project workspace management

It will be possible to save GIS project workspace into the SDI database and to share them under the predefined grant policy.

10 Security aspects

Since there is no application over the SDI database and several client software could be used to directly access to data sources, user groups and user roles will be stored at DBMS level in order to automatically apply grants and permissions to the items that will be created inside it.

Any client software connected to the SDI must **avoid storing credentials in clear mode** in order to prevent account theft by hacking datasets and project files. As an example, using QGIS accessing PostGIS it is strongly recommended NOT to use the “Base authentication” method to configure the connection because this may cause the theft of credentials by malicious people who can read them in any QGIS project files (.QGZ / .QGS) sent by email or in other ways.