

LEarning and action alliances for NexuS EnvironmentS  
in an uncertain future

# LENSES

## WP7

### D7.1 Data Integration and Visualisation: strategy and results

Eleni Ntzioni, Efstathia Chatzitheodorou, Christina  
Papadaskalopoulou, Dimitris Tassopoulos (DRAXIS)

10/04/2023

#### Project coordinator

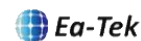


#### Project partners



#### Project Website

[www.lenses-prima.eu](http://www.lenses-prima.eu)





## Deliverable 7.1: Data Integration and Visualization: strategy and results



<b>Project no.</b>	2041
<b>Project acronym:</b>	LENSES
<b>Project title:</b>	Learning and action alliances for NEXUS environments in an uncertain future
<b>Call:</b>	PRIMA call Section 1 – Nexus 2020, Topic 1.4.1-2020 (IA).
<b>Start date of project:</b>	01.05.2021
<b>Duration:</b>	36 months
<b>Deliverable title:</b>	D7.1 Data Integration and Visualisation: strategy and results_First version
<b>Due date of deliverable:</b>	December 2021
<b>Project Coordinator:</b>	Stefano Fabiani, Council for Agricultural Research and Economics (CREA)
<b>Organisation name of lead contractor for this deliverable:</b>	DRAXIS ENVIRONMENTAL S.A.
<b>Lead Authors</b>	Eleni Ntzioni (DRAXIS), Efstathia Chatzitheodorou (DRAXIS)
<b>Email</b>	entzioni@draxis.gr
<b>Contributions from</b>	Giorgos Letsos, Spyros Papagrighoriou, Christina Papadaskalopoulou
<b>Internal Reviewer 1</b>	Stefano Fabiani (CREA)

Dissemination level		
PU	Public	
CO	Confidential, restricted under conditions set out in Model Grant Agreement	CO
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

History			
Version	Date	Reason	Revised by
0.1	27/10/2021	TOC	Eleni Ntzioni
0.3	05/11/2021	First draft	Eleni Ntzioni, Efstathia Chatzitheodorou
0.5	26/11/2021	Second draft	Eleni Ntzioni, Spyros Papagrighoriou
0.8	10/12/2021	Third draft	Giorgos Letsos, Efstathia Chatzitheodorou
1.0	20/12/2021	Prefinal	Eleni Ntzioni, Christina Papadaskalopoulou
2.0	21/12/2021	Final	Eleni Ntzioni, Stefano Fabiani
3.0	10/04/2023	Mid-Term Review revision	Christina Papadaskalopoulou, Dimitris Tassopoulos

## Executive summary

LENSES aims at enhancing system understanding, gathering and structuring knowledge to unravel complexity and manage uncertainty in WEF systems, with their dynamic evolution. In order to build resilient Nexus systems that recognize climatic, social, technological or environmental change, it is important to assess how actions in one area may affect the coordination of institutions and policies for resource management in another area. For this reason, Participatory SDMs will be developed to exploit the potential of systems thinking based on collective intelligence as a suitable and effective philosophy to support building a systemic approach to the Nexus. PSDM will benefit from the integration of highly specific models and data (e.g., hydrological models, climate data, environmental economics, etc.) provided by the LENSES Observatory.

This document presents the first version of deliverable 7.1 entitled *"Data Integration and Visualization: strategy and results"*, and it provides the approach for the implementation of the 1<sup>st</sup> version of the LENSES Observatory, a data repository storing all data collected from the pilots and make them accessible to the LENSES project partners. The LENSES Observatory will allow the collection of data from the pool of available sources, internal and external to the project. Different alternatives regarding the types of the incoming data are assessed, in order for them to be optimally integrated and maintained into state-of-the-art, distributed and scalable databases, so as to be able to uphold large-scale querying and complex analysis. This document reflects the initial strategy and results of the data integration and visualization, as well as the architecture of the system and the design and implementation of the LENSES Observatory.

## Tables of Contents

List of figures .....	6
List of tables.....	6
1. Introduction.....	7
2. Overall approach .....	8
3.1 Actions and timeline.....	8
2.2 Dependencies .....	9
2.3 Data collection protocol .....	10
2.4 Design, development and evaluation of Observatory.....	12
3. Data collection – initial version .....	13
4. Data repository – initial version .....	15
5.1 System architecture.....	15
5.2 Main features of the Observatory .....	17
5.2.1 Welcome page.....	17
5.2.2 Pilots .....	19
5.2.3 Sectors .....	21
5.2.4 Search for datasets .....	23
5.2.5 Dataset.....	25
5.2.6 Register and login .....	27
6. LENSES Observatory expandability.....	29
7. Conclusions.....	31

## List of figures

Figure 1: Actions and timeline for LENSES Observatory implementation.....	9
Figure 2: High level depiction of LENSES data flows .....	9
Figure 3. Interconnections of Observatory with other tasks of the LENSES project.....	10
Figure 4. LENSES Observatory system architecture .....	15
Figure 5: Welcome Page of the LENSES Observatory.....	18
Figure 6: Landing page after Login .....	19
Figure 7: Pilots Page .....	20
Figure 8: Page of Doñana National Park pilot.....	21
Figure 9: Sectors Page .....	22
Figure 10: Page of Climate projections sector.....	23
Figure 11: Datasets Page .....	24
Figure 12: Datasets in the pilot's page with applied filters .....	25
Figure 13: Dataset overview page .....	26
Figure 14: Download Data and Resources.....	27
Figure 15: Pilot’s page in the editing mode, tab “Members” .....	28

## List of tables

Table 1: Actions and timeline for Observatory implementation.....	8
Table 2: Template for dataset description .....	11
Table 3: Data included in the 1 <sup>st</sup> version of the Observatory.....	13

## Introduction

The goal of LENSES is to contribute to improved water allocation and enhanced food security while preserving ecosystems and aiding climate change adaptation, by supporting the operationalization of the Nexus paradigm (from Nexus Thinking to Nexus Doing). Stakeholders should be provided with the suitable tools in order to collectively create the right mechanisms to analyze and manage trade-offs and synergies among the nexus domains. The development of the LENSES Observatory assists on providing all necessary data to the LENSES project partners, in the form of a platform with an interactive interface and a digital pool of several types of heterogeneous datasets, collected from a wide range of sources. Data are gathered from different pilot areas, regarding different sectors of the Nexus analysis. They refer to both in-situ and remote sensing bio-geophysical data and indicator products as well as socio-economic indicators relating to Nexus, and they are available in various data formats, that may be useful for the analyses taking place. Moreover, anyone interested to retrieve such data will be able to do so through a user-friendly and easy-to-use interface, which also allows for dataset visualization, before selecting the more suitable one. This document describes the approach for the 1<sup>st</sup> version of the repository and the data presented there; thus, it will get updated, as this is an ongoing procedure with a continuous collection of data.

In **Section 3** of the document, the overall approach for the data collection is presented, along with specific actions and a timeline. Additionally, the dependencies with other tasks, the expected outputs, as well as the data collection protocol are described. Lastly, the design, development and evaluation process of the LENSES Observatory are mentioned.

In **Section 4**, there is a thorough description of the data collected for the 1<sup>st</sup> version of the LENSES Observatory.

In **Section 5**, the Observatory is described in detail. More specifically, details regarding the system architecture of the 1<sup>st</sup> version are presented, as well as the main features of the repository and the functionalities, accompanied by screenshots of the current implementation of the LENSES Observatory.

In **Section 6**, the document concludes with the Observatory expandability aspects and the next steps that are expected to be taken for future versions.

## 1. Overall approach

In the context of T7.1 the implementation of the LENSES Observatory is envisaged and more specifically a repository that will store and serve all data produced within LENSES. In order to develop the Observatory platform, many intermediate actions need to be defined and implemented, regarding both the data collection and the Observatory development. In this chapter, all actions taken so far, along with the planned next steps, are presented. Furthermore, the identified dependencies of the Observatory, as well as the approach for the Data collection protocol are discussed. The methodology for the development of the Observatory is introduced and will then be further analyzed in section 5.

### 3.1 Actions and timeline

Creating the LENSES Observatory is, and will continue to be, an iterative process, which involves many actors and several steps, as presented below. This process mainly includes the definition of the data collection protocol and rounds of data collection and implementation or enhancements of the Observatory. Table 1 and Figure 1 display the whole conceptualized process and its respective timeline up to M18 of the project.

*Table 1: Actions and timeline for Observatory implementation*

No.	Action	Timeline
1	Identify interconnections with other WPs and Tasks, inputs and outputs.	M3
2	Define 1 <sup>st</sup> version of the Data collection protocol.	M4
3	Identify data that can be included in the 1 <sup>st</sup> version of the Observatory.	M6
4	Implement the LENSES Observatory 1 <sup>st</sup> version.	M7
5	Finalize pilot data that can be part of the 1 <sup>st</sup> version and collect them along with metadata.	M7
6	1 <sup>st</sup> version of LENSES Observatory available with sample data from pilots and WP7 outputs	M8
7	Evaluation of Observatory 1 <sup>st</sup> version and user feedback.	M12
8	Continue updating the Observatory as new data become available and T8.2 progresses.	M12
9	Finalize visualization requirements in collaboration with end users.	M14
10	Have an updated version of the Observatory with all available data (up to that point) and some initial visualizations.	M18

As shown in the table above, the first step was to identify the interconnections with other work packages and tasks, as described in Section 3.2. Then, a first version of the Data collection protocol was prepared, as presented in Section 3.3. The identification of data that can be part of the initial version of the Observatory and the implementation of the tool were parallel processes, resulting in a complete 1<sup>st</sup> version that is described in Section 5. The final data included in the initial version are presented in Section 4.



Regarding the next steps, updating the Observatory with data will be a continuous process and will involve the collaboration with many project tasks and partners. Due to dependencies with other tasks, defining further functionalities of the Observatory, such as visualization approaches, will be one of the next actions envisaged in the following months. The next version of the Observatory is expected to be implemented by October 2022, including a substantial number of datasets and respective visualizations.

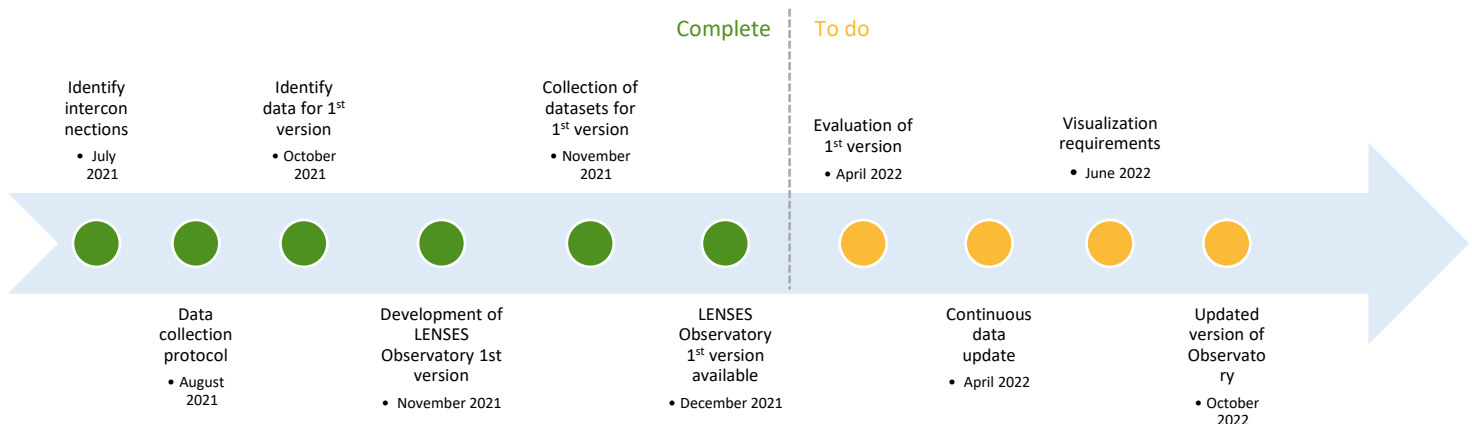


Figure 1: Actions and timeline for LENSES Observatory implementation

## 1.2 Dependencies

For the purposes of the LENSES Observatory, a large amount of heterogenous data will be collected from the LENSES pilots in a single platform, in order for them to be accessible from all partners and to be used in various tasks, as necessary. Moreover, the Observatory will store the results of WP7 tasks, that may be useful to other work packages, thus creating a single point of reference for the data used and produced in the project.

As seen in Figure 2, the high-level conceptualization of the data flows in the project includes input in the form of pilot data that will be used directly in WP7 models as well as in WP4 Participatory System Dynamics Modelling (PSDM) analysis. The models developed in the context of WP7 will produce certain results in the form of indicators, that will be used, in turn, in other WP7 tasks and in WP4 PSDM analysis. The main resulting data products that are expected to be stored and provided through the Observatory are the Pilot data and the results of WP7 models.

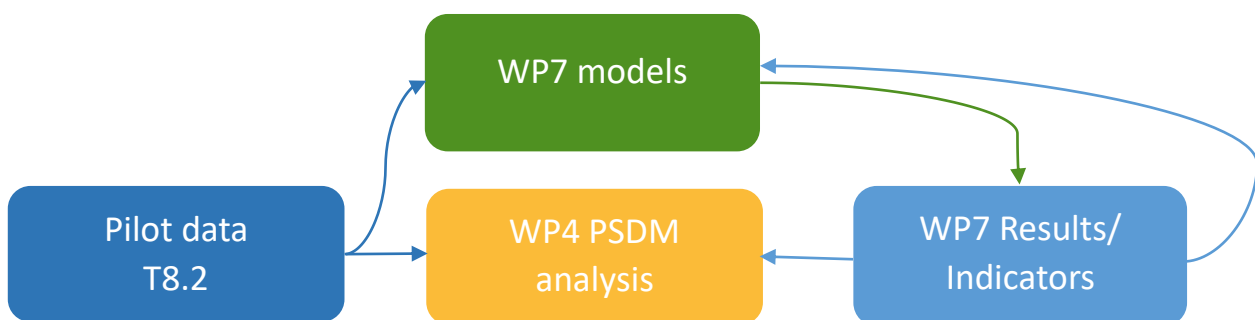


Figure 2: High level depiction of LENSES data flows

Figure 3 shows the dependencies in more detail, this time per task, rather than work package in general. For each task, the direction of the arrow denotes the data flow, i.e., whether this task provides input to the Observatory, whether it receives data, or both. The tasks associated to the Observatory in respect to data exchange are the following:

- 4.2 Participatory System Dynamics Modelling,
- 4.3 PSDM for scenario analysis,
- 7.2 Climate Projections & Climate Risk Assessments,
- 7.3 Land-use mapping & EO-based indicators,
- 7.4 Water accounting, allocation and planning.

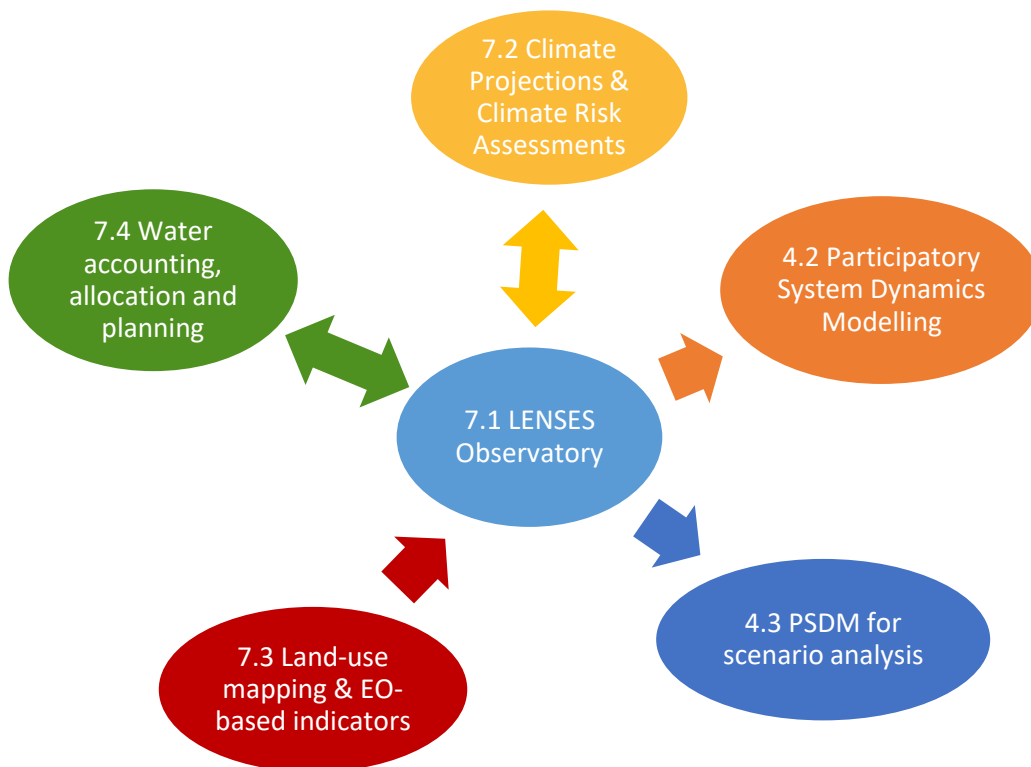


Figure 3. Interconnections of Observatory with other tasks of the LENSES project

In order to collect the data regarding the LENSES pilots, activities to identify the specific data requirements are necessary, led by task 8.2. Based on the information that will be documented in D8.2 “LENSES pilots’ data generation”, the data collection will take place as described in the data collection protocol and the datasets will be included in the LENSES Observatory. As the task 8.2 is currently in the process of identifying the data requirements, for the 1<sup>st</sup> version of the Observatory an indicative number of datasets are included, to showcase the approach and the functionality.

## 1.3 Data collection protocol

Since the LENSES project includes seven pilots with their own diverse data sources and on top of that many other tasks producing data products that will be included in the repository, a Data collection protocol is essential, to define a unified approach for the data collection process. At the current, early stages of the

project, the activities of data identification in WP8 are not finalized, thus there is limited information on the data that is expected to be collected in the LENSES Observatory. For that reason, a 1<sup>st</sup> version of the Data collection protocol is prepared, taking into account the information gathered so far, and it will be updated during the course of the project to support the data requirements and produced data products.

The main aspect of the Data collection protocol is the Data catalogue, i.e., a centralized collection of data descriptions, to make it easier to retrieve, manage and use them. This Data Catalogue consists of two parts: (i) Dataset description template and (ii) Dataset summary table. The Dataset description template contains both descriptive information that can serve as the metadata of each dataset, as well as descriptions about the type of data and how this data can be retrieved. This information is required by the technical partners of the project so as to be able to develop the data collection mechanisms. The Dataset description template contains instructions for the data details requested, so that all the datasets are described in a unified manner. The Dataset description template is presented in Table 2 along with a short explanation of each field.

Table 2: Template for dataset description

Field	Description
<b>Title of the dataset</b>	Characteristic short title of the dataset
<b>Pilot</b>	One of: “Doñana National Park”, “Galilee, Hula Valley”, “Gediz Basin and Delta”, “Koiliaris Critical Zone Observatory”, “Middle Jordan Valley, Deir Alla”, “Pinios River Basin” and “Tarquinia Plain”
<b>Sector</b>	One of: Water/ EO indicators/ Land use/ Climate projections/ Climate risk/ Pilot data
<b>Tags</b>	Keywords describing the dataset
<b>Dataset provider</b>	Organization/project that developed/maintains/provides the dataset e.g., a municipality, a private company, your organization etc.
<b>Dataset description</b>	Description of what type of data is included in the dataset in terms of (i) content, (ii) parameters and units of measurement, (iii) geographical coverage, (iv) acquisition means e.g., sensors, satellite data, surveys, models etc.
<b>Dataset timespan</b>	Does the dataset contain time series data? If yes, what is the first and last date available? If not, N/A.
<b>Time interval</b>	If the dataset contains time series data, what is the time interval between values? If not, N/A. e.g., 5 min, daily, monthly, yearly etc.
<b>Data update frequency</b>	Is the dataset static (created once and is complete) or is it updated frequently? How often are new data available in the dataset? e.g., every month, every year or N/A
<b>Dataset access details</b>	Data type e.g., excel file (or other format), how it will be provided e.g., provided through communications (emails, FTP server), file available in a private server that requires access, or other option
<b>Other technical info</b>	Additional information you may want to include, relating to accessing the dataset, if applicable. If not, N/A.

The Dataset summary table provides an overview of the data sources used for the 1<sup>st</sup> version of the LENSES Observatory and serves as a reference. The data details for all datasets managed so far are recorded in a harmonized way in this table. The Data summary table will be available in the One drive space of the LENSES project, so that all interested consortium partners can consult it at any time.

## 1.4 Design, development and evaluation of Observatory

The aim of the LENSES Observatory is to provide a data repository that will hold all data collected for the LENSES project. This repository should allow for data harmonization and visualization, to allow for drawing relationships among Nexus components. The first step towards the design of the repository is to define the basic needs it should address regarding the availability of information and its use within the project. These needs translate into the requirements included in .

No.	Requirement
1	The Observatory should allow for data research through keywords or free text.
2	The Observatory should allow for use of filters.
3	The Observatory should provide search results in a list, as a catalogue of available datasets.
4	The Observatory should include detailed descriptions of provided datasets (relevant metadata).
5	The Observatory should allow for previewing the dataset before download, where possible.
6	The Observatory should provide download functionality for all datasets included in the repository.
7	The Observatory should include APIs to access the data, where possible.
8	The Observatory should support user accounts and register/login functionality.

These requirements were presented to the consortium for a first validation with the partners who will take the role of the tool's users, and then they served as a basis for the design of the Observatory. For that purpose, research was implemented in order to identify open-source data management systems (DMS) that can support these functionalities and can be customized to be tailor-fitted to the project needs. Such is the case with the 1<sup>st</sup> version of the LENSES Observatory, as is described in detail in Section 5. The design entails further understanding of the data produced in the project, their possible categorizations, useful filters for the search functionality and finally using the color palette designed for the project.

The first version of the repository focuses on the data import and data provision functionalities, to facilitate data exchange. For this 1<sup>st</sup> version, a structured evaluation will take place, to gather feedback from the users, validate assumptions made, identify weak points, examine if their needs are met and receive suggestions for enhancements that can be implemented in future versions of the repository.

Concerning future versions of the repository, there will be an emphasis on the implementation of visualizations in the platform and a similar methodology will be used. First of all, some requirements will be set with the collaboration of the users and then the visualizations will be implemented. These may vary according to the data, from time-series graphs to interactive maps for geospatial data. In future versions, the number of available datasets and visualizations will continue to increase, as they become available in the project.

## 2. Data collection – initial version

LENSES sets out to provide all necessary data for the Nexus analysis through the Observatory, in order to produce indicators relating to Nexus for all pilot areas. For that reason, data from the five pilot areas, as well as indicators produced by the analyses of WP7 and WP4, will be part of the data products provided in the Observatory. The detailed requirements for these data products will be defined by the activities of task 8.2, but for the purposes of implementing the 1<sup>st</sup> version of the Observatory, an indicative number of datasets were identified and collected. For the collection of the datasets, the data collection protocol was followed. The template was circulated to the respective partners and some iterations followed, in order for the information to be provided in a homogenous way by all parties.

For the initial Observatory version, the **12 collected datasets** come from two main sources:

- i. Data produced by task 7.2 “*Climate Projections & Climate Risk Assessments*”, including climate projections for the “Doñana National Park” pilot area.
- ii. Data identified in the context of task 8.2 “*Assembly and deployment of LENSES solution elements*”, regarding the “Tarquinia Plain” pilot area.

A summary of the data included in the 1<sup>st</sup> version of the Observatory is included in Table 3. It is important to mention that not all pilots are represented in the current version, rather the available data refer to two pilots: “Doñana National Park” and “Tarquinia Plain” pilot areas. There are indicative datasets from the two main categories of data products expected to be part of the Observatory, i.e., pilot data and indicators. Five out of six sectors are populated with data at the moment, and most of the data are geospatial data. All data collected, were prepared by each providing partner, and were provided through email communications; no datasets were collected from external sources so far.

Table 3: Data included in the 1<sup>st</sup> version of the Observatory

Field	Description	Data collected- available info
<b>Title of the dataset</b>	Datasets collected	<ul style="list-style-type: none"> <li>▪ Climate projections – Temperature</li> <li>▪ Climate projections – Precipitation</li> <li>▪ Climate projections - Potential Evapotranspiration</li> <li>▪ Doñana study area</li> <li>▪ Climate data</li> <li>▪ Tarquinia river basin study area</li> <li>▪ Demographic data</li> <li>▪ Land use/cover data 1/25,000</li> <li>▪ Land use/cover data (detailed)</li> <li>▪ Soil map 1/250,000</li> <li>▪ Irrigated area</li> <li>▪ Territorial economic data</li> </ul>
<b>Pilot</b>	Pilots represented	<ul style="list-style-type: none"> <li>▪ Doñana National Park</li> <li>▪ Tarquinia Plain</li> </ul>
<b>Sector</b>	Sectors represented	<ul style="list-style-type: none"> <li>▪ Pilot data</li> <li>▪ Water</li> <li>▪ Land use</li> <li>▪ Climate projections</li> <li>▪ Climate risk</li> </ul>

<b>Tags</b>	Tags relevant to the current datasets	<ul style="list-style-type: none"> <li>▪ Doñana, Tarquinia area</li> <li>▪ Climate, Scenarios, Precipitation, Temperature, Evapotranspiration, RCP4.5, RCP8.5</li> <li>▪ Population data, corine land cover map, land use map, soil map, Agricultural irrigated area, Economic data</li> </ul>
<b>Dataset provider</b>	Organizations providing the current datasets	<ul style="list-style-type: none"> <li>▪ DRAXIS S.A.</li> <li>▪ AGRISAT</li> <li>▪ CREA</li> <li>▪ Water User Association (WUA)</li> </ul>
<b>Data acquisition means</b>	Original sources of input data	<ul style="list-style-type: none"> <li>▪ National Statistical System (ISTAT- <a href="http://www.istat.it">www.istat.it</a>) database</li> <li>▪ Retrieved from Copernicus Climate Data Store (CDS)</li> <li>▪ Retrieved from Earth System Grid Federation (ESGF)</li> <li>▪ Prepared by CREA and Agrisat</li> <li>▪ Corine Land Cover</li> </ul>
<b>Data timespan</b>	Timespan overview of current datasets	<ul style="list-style-type: none"> <li>▪ 1986-2005</li> <li>▪ 2031-2090</li> <li>▪ 2021</li> <li>▪ 2000-2020</li> <li>▪ 2012</li> <li>▪ 2015</li> </ul>
<b>Time interval</b>	Overview of time between the measurements of time series data	<ul style="list-style-type: none"> <li>▪ Annual mean</li> <li>▪ Annual with daily mean</li> <li>▪ N/A</li> </ul>
<b>Data update frequency</b>	Frequency that current data are collected	<ul style="list-style-type: none"> <li>▪ N/A</li> <li>▪ Annual</li> </ul>
<b>Dataset access details</b>	Formats of the current datasets and access means	<ul style="list-style-type: none"> <li>▪ .doc</li> <li>▪ .shp</li> <li>▪ .nc</li> <li>▪ .tif</li> <li>▪ .csv</li> <li>▪ .zip</li> <li>▪ .xlsx</li>   <li>▪ Provided through communications</li> <li>▪ Provided through email</li> </ul>

### 3. Data repository – initial version

In this chapter, the system architecture of the Observatory, as well as the implemented components and their functionalities will be discussed. It is worth mentioning that the system architecture constitutes a first approach that will undergo a trial operation in order to be validated or modified as required. Respectively, the components included in this initial version of the repository will be evaluated through user feedback, thus improvements and additions are expected in future versions. Additionally, due to the fact that the data identification process is an ongoing procedure, more data will be collected during the lifetime of the project and will be presented in future versions of the repository.

### 5.1 System architecture

The design and development of the LENSES Observatory was based on the arising need of the project, to facilitate the exchange and use of data produced from different tasks and partners of the project. Firstly, research was conducted on existing open-source data management systems (DMS), that integrate tools to streamline publishing, sharing, finding and using data, or even previewing and visualizing data, to discover insights before using them. CKAN open-source software<sup>1</sup> was selected among others and the system architecture was tailored to include this DMS. The subsystems along with their characteristics are presented below.

The system architecture meets basic technical requirements such as:

- Availability: Continuous provision of services to the end user
- Extensibility: Ability to extend the architecture to support new services
- Security: Protection against risks, viruses, access breaches, publication of incorrect data
- Scaling: Ability to upgrade requirements for maximum performance
- Reliability: Accuracy and consistency of services provided
- Ease of management: Monitoring procedures to ensure quality service delivery

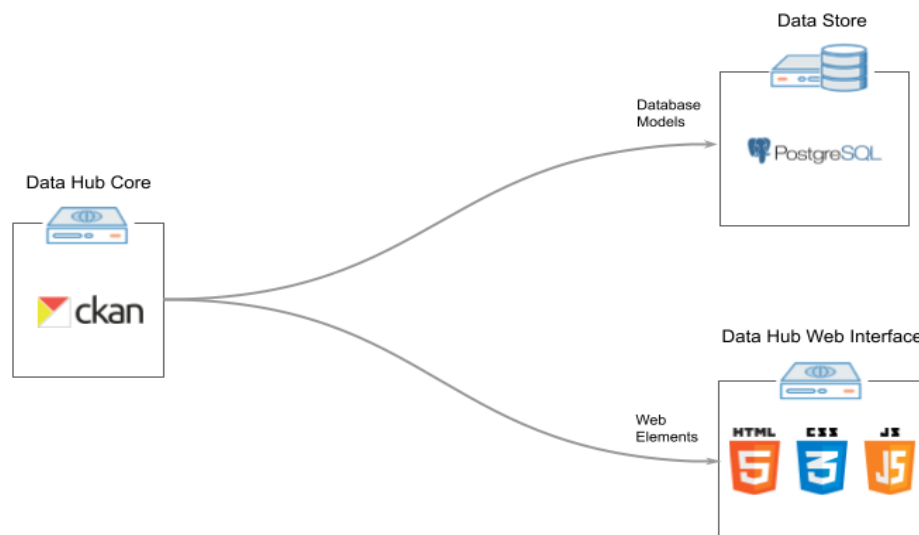


Figure 4. LENSES Observatory system architecture

<sup>1</sup> <https://ckan.org/>

The architecture consists of three main subsystems whose purpose is the implementation and support of the overall process of data import, data searching and browsing by the users and data accessing. It includes subsystems responsible for the import, processing and storage of data relevant to the seven pilots of LENSES, as well as the search and presentation of results to the end users, through a friendly online environment.

As seen in Figure 4, the core subsystem is the “Data hub core”, where the data are injected via an exposed API. Storing of the data is handled by the “Data store” and they will be provided to the users either through APIs directly to other applications or through the “Data hub Web interface” which will be the means of interaction between the system and the users.

### Data hub core:

The Data hub core subsystem is based on Python programming language and the CKAN Portal. It acts as an Orchestrator of the system, while applying the business processes to the data it draws from the peripheral subsystems.

The main functionalities of Data hub core subsystem are:

- The receipt of data coming from the crawling service through the User Interface Subsystem.
- The storage of files and geospatial data in the corresponding components of the database subsystem.
- Apply the business rules of the platform.
- Exchange data between the components of the Data handling subsystem.
- Convert the results to a user-friendly format and display them through the Interface system

A mechanism responsible for the security of the system and its contents is also implemented. Access control lists include lists of users, groups and roles that have access to the content or interact with components of the platform. This unit also shields the application from third-party attacks such as SQL injection, cross-site scripting (XSS), cross-site request forgery (CSRF) and cookie hijacking.

### Data Store:

The Data Store Component is used in order to store all the data of the system. All the data types are stored in a modeled database format. The types of data that refer to the business data, defined as the data of users and generally data related to the business operation of the system, are better handled by an object-oriented database, such as PostgreSQL, which offers a variety of powerful index types to best match a given query workload. Moreover, it offers performance optimizations such as parallelization of read queries, table partitioning, and just-in-time compilation of expressions. Besides these characteristics, a really important factor for choosing PostgreSQL is that it has a spatial database extender, PostGIS, which supports geographic objects and allows location queries in geographical geometries. Since the majority of the data is expected to be geospatial, querying a PostGIS database allows more powerful and precise insights.

### Data hub Web Interface:

The User Interface (UI) is the human-computer point of interaction, where the graphical user interface (GUI) of the web application is developed. The web interface is based on HTML and JavaScript, in combination with Jinja, a modern and designer-friendly templating language for Python, embedded by CKAN. Jinja is fast, expressive and extensible. Special placeholders in the template allow writing code similar to Python syntax and then the template is passing data to render the final document.



## 5.2 Main features of the Observatory

**Dataset:** The data publishing unit in the LENSES observatory is called “dataset”, meaning a parcel of data. For example, it could be the temperature values for a specific location at a specific time. A dataset consists of “metadata” and a number of “resources”, which hold the data itself. The data formats can vary, from CSV and Excel files, to images and linked data in RDF format. Moreover, a dataset can contain a large or a small number of resources.

**Metadata:** A loose term of metadata is “data about data”. Metadata identify and describe all aspects of a specific dataset (i.e., the who, why, what, when and where) that allow the understanding of the physical format, content as well as context of the data. The metadata accompany the Datasets and are provided through the Data collection protocol.

**Pilots:** The datasets of the project, that are included in the Observatory, are related to one of the pilots of the LENSES project, i.e., the “Doñana National Park”, “Galilee, Hula Valley”, “Gediz Basin and Delta”, “Koiliaris Critical Zone Observatory”, “Middle Jordan Valley, Deir Alla”, “Pinos River Basin” and “Tarquinia Plain”. In the first version of the observatory, indicative datasets from all pilots are successfully collected.

**Sectors:** In order to further categorize the datasets in the Observatory, several sectors are introduced. A dataset can belong to more than one sectors. For the first version of the repository, several sectors have been identified and 5 are populated with data: Water, Climate projections, Land use, Climate risk and Pilot data. These sectors represent a simple way to help users search and access data thematically.

**Administrator:** There is an administrator for the repository that is responsible for managing the content and the users, as well as assigning authorization rights to users.

**User:** The repository can be accessed only from the partners of the LENSES project. These users can view the uploaded datasets and access them through the interface.

### 5.2.1 Welcome page

The welcome page of the LENSES Observatory introduces the user to the Observatory and allows the user to login with their credentials, in order to access the datasets. The welcome page contains a small description of the project and the Observatory, along with statistics showing (i) the numbers of the pilots, (ii) the available datasets and the (iii) sectors provided in the tool (Figure 5).

When the login is complete, the user will be redirected to their “Dashboard”, where they will be able to see activity from items they are following, i.e., Pilots, Sectors and Datasets. At the top of the page, there is a menu for allowing the user to easily navigate to the Datasets page, the Pilots page, the Sectors page and the About page (Figure 6).

the  
LENSES  
project

## Welcome to the LENSES Observatory

Log In

The objective of LENSES project is to enhance system understanding, gathering and structuring knowledge in order to unravel complexity and manage uncertainty in Water, Ecosystem, Food systems, with their dynamic evolution.

The LENSES Observatory serves as a treasure chest of data, to facilitate information exchange internally. This data repository also supports data preview and visualizations, to allow for drawing relationships among Nexus components. The Observatory will continue to be populated with more datasets, as they are being produced in the LENSES project.

### LET'S TALK WITH NUMBERS

<b>7 PILOTS</b>	<b>12 DATASETS</b>	<b>6 SECTORS</b>

**Funded by the European Union**

This project has received funding from the European Union's Horizon 2020 research and innovation programme, specifically the Partnership for Research and Innovation in the Mediterranean Area (PRIMA Foundation), under grant agreement No 2041.

[Visit Lenses website to learn more about the project!](#)

Figure 5: Welcome Page of the LENSES Observatory

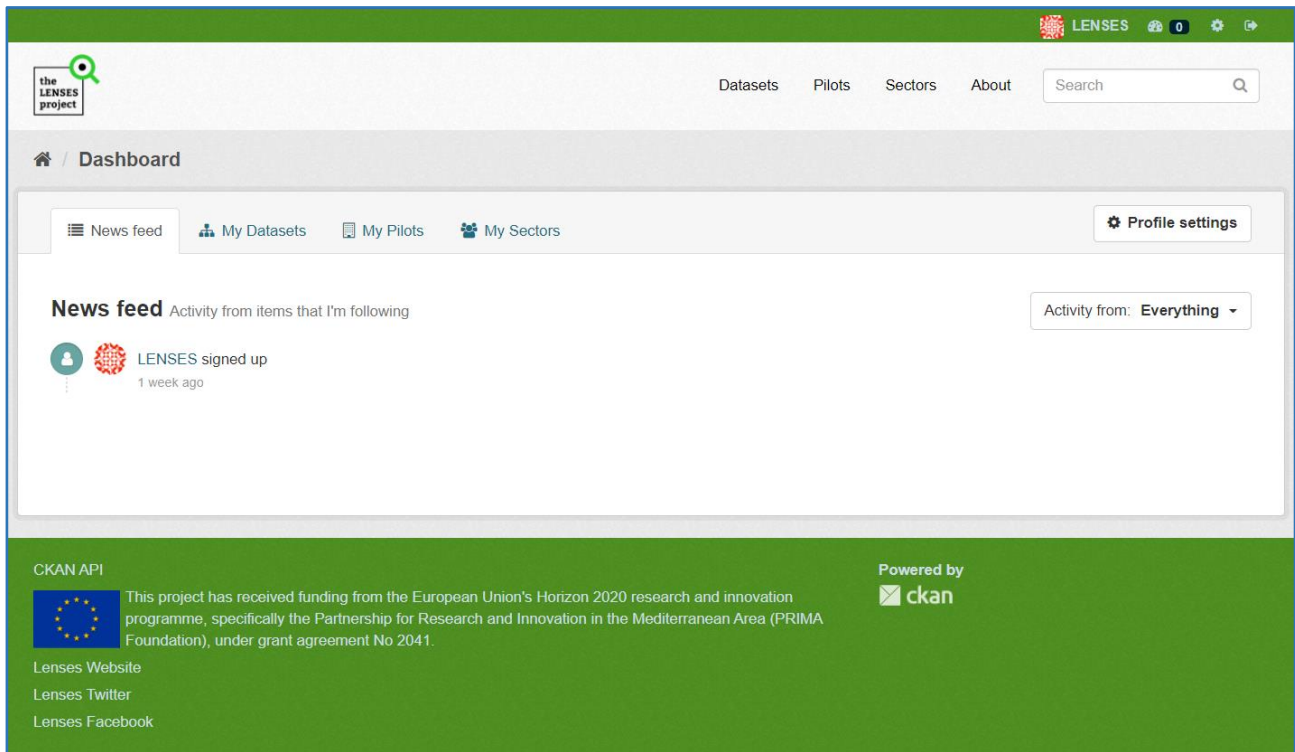


Figure 6: Landing page after Login

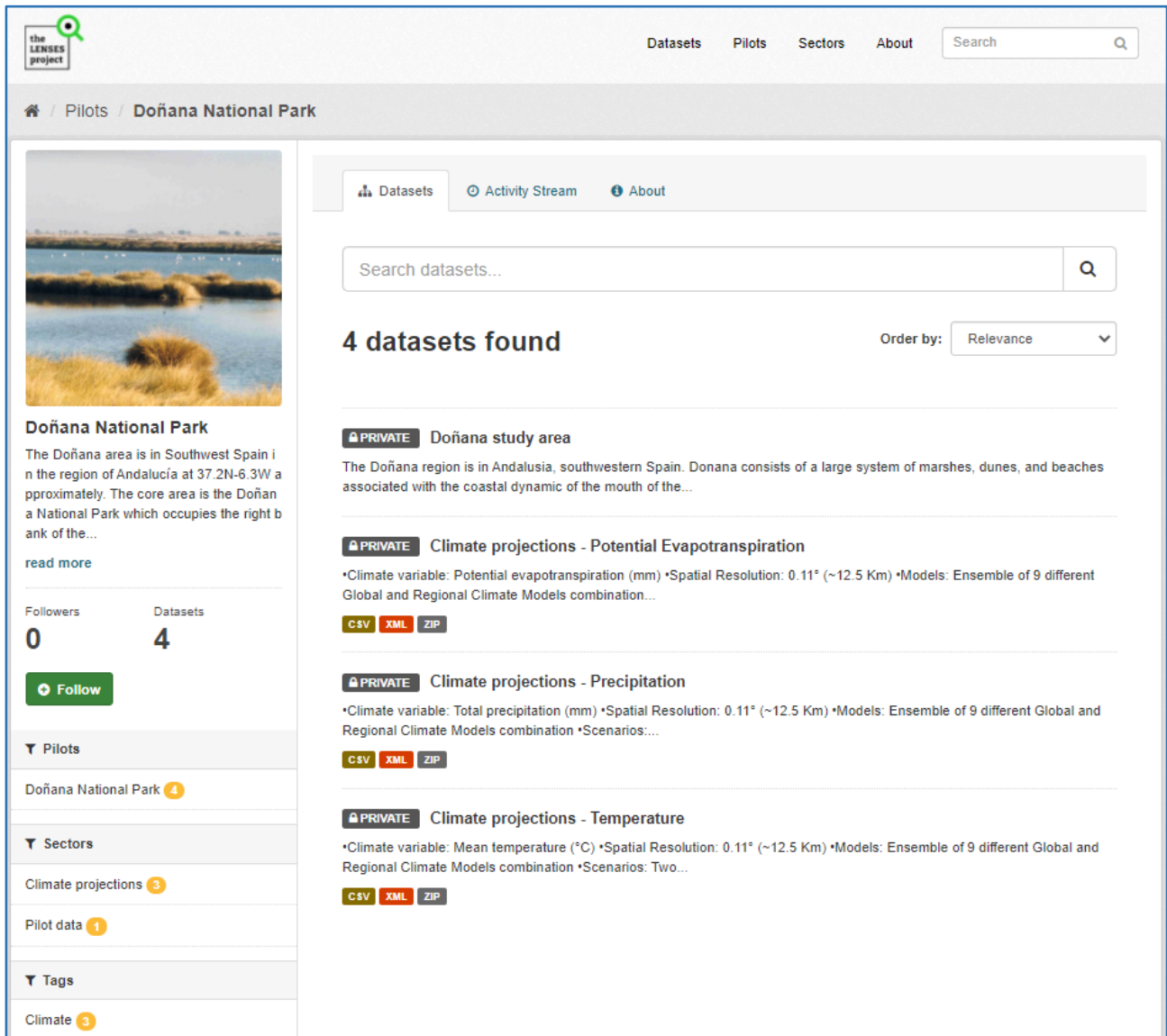
## 5.2.2 Pilots

The user can navigate to the Pilots page by selecting “Pilots” from the menu bar (Figure 7). Each one of the pilots has a page with information regarding this particular pilot (tab “About”), and the user can search within the datasets (tab “Datasets”) and see the latest activities related to the datasets of that specific pilot (tab “Activity Stream”).

The screenshot shows the 'Pilots' page on the LENSES website. At the top, there is a navigation bar with 'Datasets', 'Pilots' (highlighted), 'Sectors', and 'About', along with a search bar. Below the navigation, the page title is 'Pilots'. On the left, a sidebar contains a section titled 'What are Lenses Pilots?' with a brief description of the LENSES project's goals. The main content area features a search bar for pilots, a dropdown menu set to 'Name Ascending', and a grid of 7 pilot cards. Each card includes a representative image, the pilot's name, a short description, and the number of datasets available for that pilot.

Pilot Name	Description	Datasets
Doñana National Park	The Doñana area is in Southwest Spain in the region of Andalucía at...	4 Datasets
Galilee, Hula Valley	The Galilee region is the far most northern periphery of Israel (see map)....	0 Datasets
Gediz Basin and Delta	The area of 17500 km <sup>2</sup> makes the Gediz River Basin one of the most important...	0 Datasets
Koiliaris Critical Zone Observatory	The Koiliaris River watershed is a Critical Zone Observatory (CZO)...	0 Datasets
Middle Jordan Valley, Deir Alla	The Jordan Valley extends from Lake Tiberias, at an elevation of -212 meters...	0 Datasets
Pinios River Basin	The Pinios River Basin (PRB) is in central Greece, and it is one of the most...	0 Datasets
Tarquiniia Plain	Piana di Tarquinia is located in central Italy, Lazio Region, around 90 km...	8 Datasets

Figure 7: Pilots Page



The screenshot displays the LENSES Observatory interface for the Doñana National Park pilot. At the top, there is a navigation bar with 'Datasets', 'Pilots', 'Sectors', and 'About' menus, along with a search bar. The main content area is titled 'Doñana National Park' and features a sidebar on the left with a profile picture, a description, and a 'Follow' button. The main area shows a search for datasets, resulting in 4 datasets found. The datasets listed are:

- Doñana study area** (PRIVATE): The Doñana region is in Andalusia, southwestern Spain. Donana consists of a large system of marshes, dunes, and beaches associated with the coastal dynamic of the mouth of the...
- Climate projections - Potential Evapotranspiration** (PRIVATE): Climate variable: Potential evapotranspiration (mm) •Spatial Resolution: 0.11° (~12.5 Km) •Models: Ensemble of 9 different Global and Regional Climate Models combination...
- Climate projections - Precipitation** (PRIVATE): Climate variable: Total precipitation (mm) •Spatial Resolution: 0.11° (~12.5 Km) •Models: Ensemble of 9 different Global and Regional Climate Models combination •Scenarios:...
- Climate projections - Temperature** (PRIVATE): Climate variable: Mean temperature (°C) •Spatial Resolution: 0.11° (~12.5 Km) •Models: Ensemble of 9 different Global and Regional Climate Models combination •Scenarios: Two...

Figure 8: Page of Doñana National Park pilot

### 5.2.3 Sectors

In the LENSES Observatory the user is able to find various sectors that are related to the work of WP7 tasks in respect to the pilots, by selecting “Sectors” from the menu bar (Figure 9). Each sector has a dedicated page where users can search within its datasets (tab “Datasets”) and see the latest activities related to the datasets of that specific sector (tab “Activity Stream”).

The screenshot shows the 'Sectors' page of the Lenses project. At the top, there is a navigation bar with 'Datasets', 'Pilots', 'Sectors' (highlighted), and 'About'. A search bar is located on the right. Below the navigation bar, the page title is 'Sectors'. On the left, a sidebar contains a section titled 'What are Lenses Sectors?' with the text: 'To further categorize the datasets thematically in the Observatory, several sectors are introduced. A dataset can belong to more than one sectors.' The main content area features a search bar labeled 'Search sectors...' and a dropdown menu for 'Order by: Name Ascending'. Below this, a grid displays six sectors, each with a representative image and a title: 'Climate projections' (mountain landscape), 'Climate risk' (cracked earth), 'EO indicators' (satellite over Earth), 'Land use' (agricultural fields), 'Pilot data' (map with markers), and 'Water' (underwater scene).

Figure 9: Sectors Page

The screenshot displays the LENSES project website interface. At the top, there is a navigation bar with 'Datasets', 'Pilots', 'Sectors', and 'About' links, along with a search box. The main content area is titled 'Pilot data' and features a map of Italy with location markers. Below the map, there are statistics for 'Followers' (0) and 'Datasets' (6), with a 'Follow' button. A sidebar on the left lists various categories: Pilots (Tarquinia Plain 5, Doñana National Park 1), Sectors (Pilot data 6, Climate risk 1), Tags (Climate 1, Doñana 1, Economic data 1, Population data 1, Soil map 1, Tarquinia area 1, Temperature 1), and Formats (XLSX 2). The main content area shows a search bar with 'Search datasets...' and a search icon. Below the search bar, it indicates '6 datasets found' and an 'Order by: Relevance' dropdown menu. The search results list six datasets, each with a 'PRIVATE' label and a brief description:
 

- Territorial economic data**: The farm accountability data network (FADN) monitors farms' income and business activities. Farm economy data were divided by sector: crop farms, livestock farms and mixed crop and...
- Soil map 1/250,000**: Soil map called "Carta dei Suoli del Lazio" at scale 1:250.000 with relate DB are elaborated by CREA. The digital soil maps are elaborated on the base of punctual soil data...
- Demographic data**: Demographic data from National Statistical System (ISTAT- [www.istat.it](http://www.istat.it)) database. At the moment data refers only 2021, time series is available on ISTAT website. Includes an XLSX download button.
- Tarquinia river basin study area**: Piana di Tarquinia is located in central Italy, Lazio Region. The area has a size of 27.000 ha. File is prepared by CREA. Includes a ZIP download button.
- Climate data**: Agrometeorological data: Minimum and Max air temperature at a height of 2m above the surface; Minimum and maximum 2m relative humidity; Wind speed at a height of ten meters... Includes an XLSX download button.
- Doñana study area**: The Doñana region is in Andalusia, southwestern Spain. Donana consists of a large system of marshes, dunes, and beaches associated with the coastal dynamic of the mouth of the...

Figure 10: Page of Climate projections sector

## 5.2.4 Search for datasets

In order to find a specific dataset, the user can type any combination of words (e.g., “demographic”, “precipitation”, etc.) in the search box of any page. As a result, the Observatory will return a list of all corresponding search results (Figure 11).

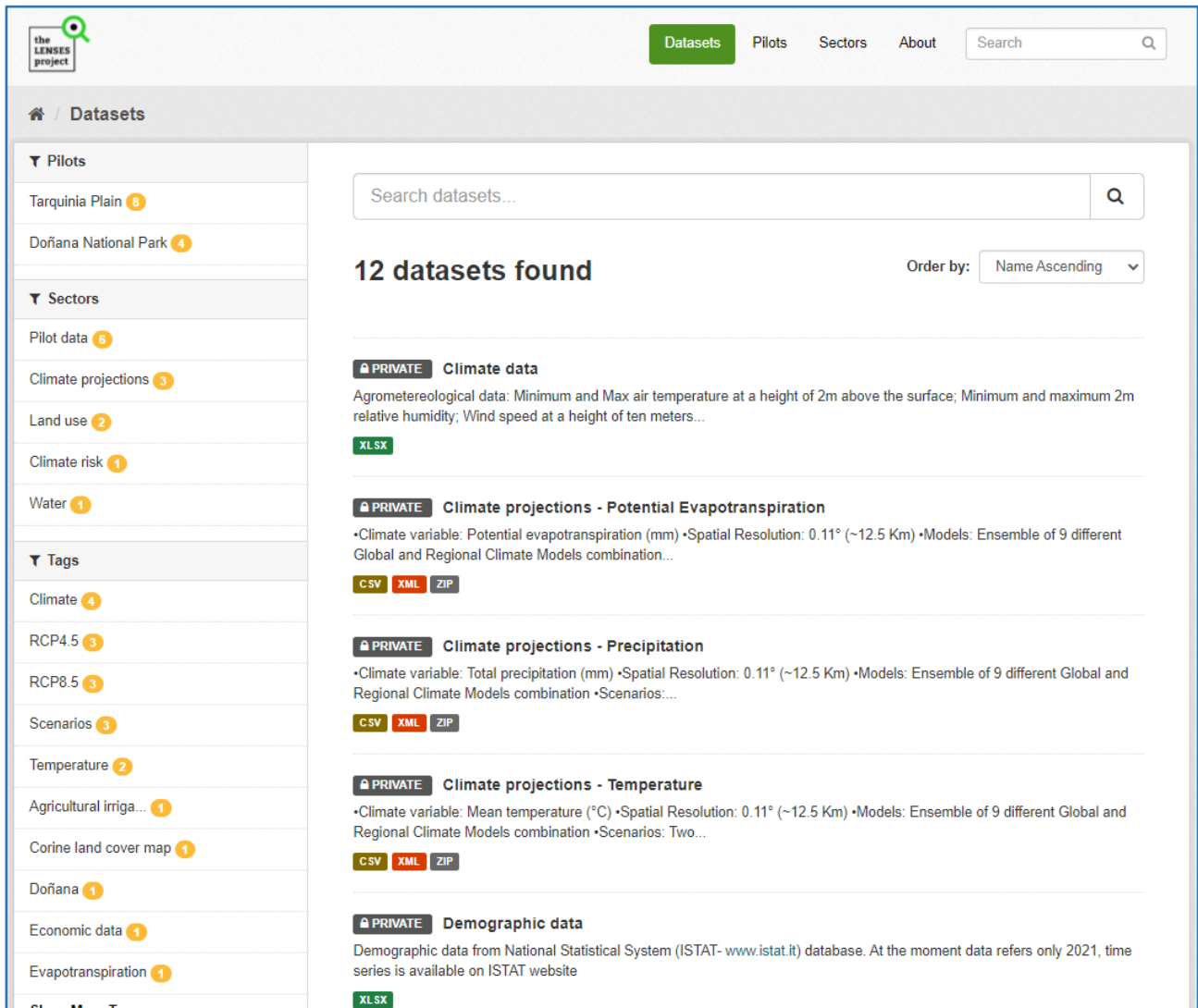


Figure 11: Datasets Page

On the search result page, the user can sort the results by relevance, name, modification date or popularity, by selecting “Order by”. They can also use the filters on the left column (Pilots, Sectors, Tags and Formats) in order to limit the results. The user can also combine filters, selectively adding and removing them, and modify and repeat the search with existing filters still in place.

Additionally, the users can select “Pilots” from the menu in order to view the seven pilots and then select the one they are more interested in and be directed to that specific pilot’s page. By typing a search query in the main search box on the page, the data hub returns search results as described above but restricted to datasets from the specific pilot (Figure 12). Apart from typing in the search box, the user can explore the datasets in that specific pilot. Respectively, they can select “Sectors” from the menu and follow the same process to explore the datasets thematically.



The screenshot shows the 'Tarquinia Plain' pilot page on the LENSES project website. The page features a search bar at the top right and navigation tabs for 'Datasets', 'Activity Stream', and 'About'. Below the search bar, there is a search input field with the text 'Search datasets...'. The results section shows '2 datasets found' and an 'Order by: Relevance' dropdown menu. The datasets are filtered by 'Land use' and 'Tarquinia Plain'. The first dataset is 'Land use/cover data (detailed)' with a 'PRIVATE' label and a 'ZIP' download button. The second dataset is 'Land use/cover data 1/25,000' with a 'PRIVATE' label and a 'ZIP' download button. The left sidebar contains a header for 'Tarquinia Plain' with a description, a 'read more' link, and statistics for 'Followers' (0) and 'Datasets' (2). Below this are sections for 'Pilots', 'Sectors', and 'Tags', each with a list of items and a close button.

Figure 12: Datasets in the pilot's page with applied filters

## 5.2.5 Dataset

Once the user finds a dataset they are interested in and selects it, the CE data hub will display the dataset page (Figure 13). On the overview page of a dataset, the user will find three tabs: "Dataset", which shows the data and resources belonging to this dataset as well as additional info (metadata), "Sectors", which shows the sectors this dataset belongs to and "Activity stream", which shows the history of recent changes to the dataset. On the left part is a static column that displays the title of the dataset and the pilot it relates to.

the LENSES project

Datasets Pilots Sectors About Search

Home / Pilots / Doñana National Park / Climate projections - ...

### Climate projections - Potential Evapotranspiration

Followers: 0 [Follow](#)

Pilot

**Doñana National Park**  
The Doñana area is in Southwest Spain in the region of Andalucía at 37.2N-6.3W approximately. The core area is the Doñana National Park which occupies the right bank of the...

[read more](#)

Dataset Sectors Activity Stream

## Climate projections - Potential Evapotranspiration PRIVATE

2 views (2 recent)

- Climate variable: Potential evapotranspiration (mm)
- Spatial Resolution: 0.11° (~12.5 Km)
- Models: Ensemble of 9 different Global and Regional Climate Models combination
- Scenarios: Two Representative Concentration Pathways - RCP4.5 (intermediate) & RCP8.5 (business as usual)

Files are spatial, with the mean annual value, for the period 2031-2090 and for the reference period 1986-2005. Retrieved from Earth System Grid Federation (ESGF) <https://esgf-data.dkrz.de/search/cordex-dkrz/> and then applied further processing  
Check D3.9 Report «Fit-for-Nexus climate projections», for further information.

### Data and Resources

CSV	DONANA EVAP 45 CSV	<a href="#">Explore</a>
CSV	DONANA EVAP 85 CSV	<a href="#">Explore</a>
CSV	DONANA EVAP HIST CSV	<a href="#">Explore</a>
XML	DONANA EVAP 45 XML	<a href="#">Explore</a>
TIFF	DONANA EVAP 45 TIFF	<a href="#">Explore</a>
XML	DONANA EVAP 85 XML	<a href="#">Explore</a>
TIFF	DONANA EVAP 85 TIFF	<a href="#">Explore</a>
XML	DONANA EVAP HIST XML	<a href="#">Explore</a>
TIFF	DONANA EVAP HIST TIFF	<a href="#">Explore</a>

Climate Evapotranspiration RCP4.5 RCP8.5 Scenarios

### Additional Info

Field	Value
Last Updated	December 13, 2021, 3:48 PM (UTC+02:00)
Created	December 13, 2021, 3:47 PM (UTC+02:00)
Data timespan	1986-2005 (Reference period) and three 20-year future periods within 2031-2090
Data update frequency	N/A
Organization	DRAXIS S.A.
Time interval	Daily mean per year

Figure 13: Dataset overview page

On the “Dataset” tab the user can see all the information of the dataset including the title, the description, the list of data and resources, the keywords associated to the dataset and the additional info. The “Explore” button on the right of each resource offers the following options to the user: “More information”, which

shows the page of the resource including additional information, and “Download”, which downloads the file directly (Figure 14). The list of keywords and the additional information presents the metadata of the dataset, provided through the data collection protocol.

The screenshot displays a section titled "Data and Resources" containing a list of nine datasets. Each dataset entry includes a file icon (CSV, XML, or TIFF), the dataset name, and an "Explore" button. The dataset "DONANA EVAP 85 XML" is highlighted, and its "Explore" button is open, showing a dropdown menu with "More information" and "Download" options. At the bottom of the list, there are filter buttons for "Climate", "Evapotranspiration", "RCP4.5", "RCP8.5", and "Scenarios".

File Type	Dataset Name	Action
CSV	DONANA EVAP 45 CSV	Explore
CSV	DONANA EVAP 85 CSV	Explore
CSV	DONANA EVAP HIST CSV	Explore
XML	DONANA EVAP 45 XML	Explore
TIFF	DONANA EVAP 45 TIFF	Explore
XML	DONANA EVAP 85 XML	Explore (More information, Download)
TIFF	DONANA EVAP 85 TIFF	Explore
XML	DONANA EVAP HIST XML	Explore
TIFF	DONANA EVAP HIST TIFF	Explore

Figure 14: Download Data and Resources

## 5.2.6 Register and login

Registered users of the Observatory can search for and download datasets without any restrictions. Registered users are created by the Administrator. The required information for the registration is username, full name, e-mail address and password. After the registration, the Administrator will assign authorization rights to the user, for them to be able to view the datasets of a specific pilot and sector. The Administrator will add the new user to the seven pilots and will assign to them the “Member” role, so they can have access to the datasets of each pilot (Figure 15).

For the 1<sup>st</sup> version of the LENSES Observatory, a Registered user account is created that will be used by all consortium partners, that are interested in retrieving the data and testing out the functionality. After the evaluation of the tool and the implementation of additional functionalities, separate accounts will be created for all interested parties and the registration process will be defined as needed.

the LENSES project

Datasets Pilots Sectors About Search

Home / Pilots / Tarquinia Plain / Manage

Edit Datasets Members View

**Add Member**

**2 members**

User	Role	
DRAXIS	Admin	
LENSES	Member	

**Tarquinia Plain**  
 Piana di Tarquinia is located in central Italy, Lazio Region, around 90 km north of Rome. The main activities are tourism (Tarquinia is a UNESCO World Heritage list from 2004)...  
[read more](#)

Figure 15: Pilot's page in the editing mode, tab "Members"

## 6. LENSES Observatory expandability

Updating the LENSES Observatory is a process that will be active during the course of the project, for many months to come. Due to the fact that there are many dependencies with tasks that are still in early stages, defining all aspects of the Observatory is not possible at the current phase. However, the Observatory is conceptualized and developed in such a way, that allows future integration and expandability of data management and provision services. The Observatory is expected to be enhanced in three ways:

### 1. Import of additional datasets to the existing platform, that will require the development of data collection mechanisms.

The role of data collection is to retrieve the data from the data sources and feed it into the Observatory for storage and further processing. The data that will be identified in the course of the project may be live data, like sensor readings or feeds from social media, or static data, like environmental reports and documents. Additionally, geospatial map data could be considered as static, as it does not change very often. Similarly, there are other data sources which deliver reports generated at regular intervals, e.g., weekly, once per month/quarter, yearly. Such static data does not require immediate, near real-time processing and analysis. It can be stored on the platform, aggregated with other data sources to provide deeper insights, and made available for offline analysis. At the moment, the current approach handles static data, due to the fact that this is the type of available datasets for the 1<sup>st</sup> version. If dynamic data sources are identified, a data collection layer will be created for the Observatory that will be able to deal with bulk data, different data formats, and structured/semi-structured/unstructured data. For that purpose, automated tasks could be scheduled periodically with cron-jobs (software used for scheduling tasks to run on the server), that will trigger the crawling mechanism enabling any new data to be crawled in the Observatory. Specifically, in the case that new data exists, it will be grabbed and ingested in the system and then forwarded to the content parser which will extract the information and store it in the database.

Moreover, further external data sources may be identified, that provide static data through Application Programming Interfaces (APIs) for retrieving the data; they could be stored in other Database Management Systems or be generated as the output of some processes. If an API to retrieve the data is provided, then specific harvesting methods will be created to transfer the data to the Observatory. In cases where APIs are not provided and the data should be manually imported to the Observatory, various solutions will be examined per case, e.g., File Transfer Protocol for bulk data.

### 2. Enhancement of functionalities of the existing platform.

As described in section 3.4, basic requirements have been identified for the LENSES Observatory that will allow the users to benefit from the platform and the data provided there. Some of the envisaged functionalities are implemented in the 1<sup>st</sup> version of the Observatory and some will be developed in future versions. More specifically, future versions of the Observatory will include (i) dataset preview functionality, where the user will be able to view tables and simple graphs of the dataset to better understand its contents and also (ii) the provision of APIs to directly access the data, instead of downloading the available resources.

### 3. Development of visualizations targeted to the indicators produced by WP7 models.

The LENSES Observatory will allow for data visualization so that the users will be able to draw relationships among water accounting, land use and earth observation based indicators, as well as climate projections and climate risk, for all pilot cases. For that purpose, the results of tasks 7.2-7.4 will be analysed in order to define some visualization requirements for the data and indicators produced. Based on these requirements, specific



## Deliverable 7.1: Data Integration and Visualization: strategy and results



visualizations will be created and provided through the Observatory, in the form of maps, tables, time-series graphs, metrics, aggregated results or other formats that are required.

## 7. Conclusions

This deliverable aims to describe the process followed for the implementation of the LENSES Observatory and the data collection in the context of T7.1 “*LENSES Observatory*”. The overall approach is presented, including the identified dependencies, the data collection protocol, the collection of the first batch of data and the development of the 1<sup>st</sup> version of the platform.

The 1<sup>st</sup> version of the LENSES Observatory provides a repository with many features allowing data search and retrieval. Due to the dependencies with many other tasks of the project and due to the fact that the data identification process is in progress, the current version contains 12 datasets produced by task 7.2 and the pilot area of “Tarquinia plain”. The Observatory will continue to be populated as data become available. Even though the data collection so far included handling of static data that were shared and uploaded directly in the platform, the architecture and implementation of the Observatory is able to support integration with various data collection mechanisms, that will be developed as the need arises.

The Observatory will continue to be operational after the completion of the project for at least 5 years. During this period, maintenance activities will be carried out to ensure the platform's proper functionality and will be updated, in case new datasets are provided by the partners.

The current document constitutes a first draft version of the LENSES Observatory and it will be updated at M36 of the project, with the final version of deliverable “*Data Integration and Visualisation: strategy and results*”.



This publication reflects only the author's view and the PRIMA Foundation is not responsible for any use that may be made of the information it contains