



Project of Strategic Interest NEXTDATA

WP2.7

D2.7A – NextData General Portal and extended brokering middleware

Responsible: Stefano Nativi (CNR-IIA)

Contribution by: Paolo Mazzetti (CNR-IIA)

Table of Contents

Introduction	3
1. Brokering architecture	3
2. Discovery functionalities	4
2.1 Data Sources Browsing	4
2.2 Query Based Search	5
2.3 Results Filtering	6
2.4 Semantic Search	6
3. Access functionalities	7
3.1 Direct download	8
3.2 Mediated access	8
3.3 Data Policy	9
Annex A. Portal Functionality schematic	11
A1: Human Computer Interaction (HCI) Overview	11
A2: HCI first level	12
A3: Discovery high-lev functionalities	12
A4: Simple Discovery high-lev functionalities	12
A5: Discovery Results Evaluation functionalities	13
A6: HCI Access and simple Transform	14
A7: HCI General Functionalities	15

Introduction

The present deliverable describes the NextData general portal and extended brokering middleware enabling the NextData System of Systems Infrastructure (ND-SoS-Ina).

The portal is the graphical user interface allowing users to search and access NextData datasets. It is a general-purpose interface providing typical data portal functionalities for searching datasets based on geographic coverage, temporal extent, and keywords, and for downloading them.

The brokering middleware is the core component of the infrastructure. It allows connecting heterogeneous data sources – supporting different interfaces and delivering data with different data and metadata formats – and harmonizing them. Thanks to the brokering framework, external applications – including the portal – can access all the datasets through a homogeneous interface as if they were provided by a single data system.

The deliverable is structured according to the two macro-functionalities provided by the NextData System of Systems Infrastructure, presenting internal architecture – the middleware – and the user interaction – the portal. Section 1 describes the discovery functionalities, while Section 2 the access functionalities. In Annex A the detailed schemas for complete portal functionalities are reported.

1. Brokering architecture

The diagram in Figure 1 shows the current deployment of the NextData Brokering architecture. These are the principal components:

- **Portal:** It provides a user friendly web interface to the NextData users, offering discovery and access functionalities (made available by the Broker) in a convenient way.
- **Broker:** the middleware, which connects the Portal with the available data sources. It acts as a mediator of discovery and functionalities for data access. Indeed, the Broker is in charge of executing the (discovery and access) requests that the users generate through the Portal. Execution can involve one or more of the available Data Providers, which are accessed by the Broker using Data provider specific protocols and data models. Finally, the results are harmonized by the Broker making them homogeneously accessible and sent back to the Portal, to be displayed to the users in a convenient way.
- **Data Providers:** Datasets are published in NextData through a set of data publishing services for the web. Different partners are responsible for these components, and each service may use its own protocol and data model for serving the data:
 - CMCC (THREDDS Data Server)
 - ICTP (THREDDS Data Server)
 - ENEA (THREDDS Data Server)
 - ISAC (THREDDS Data Server)
 - INGV (THREDDS Data Server)
 - GeoNetWork, Web Map Service

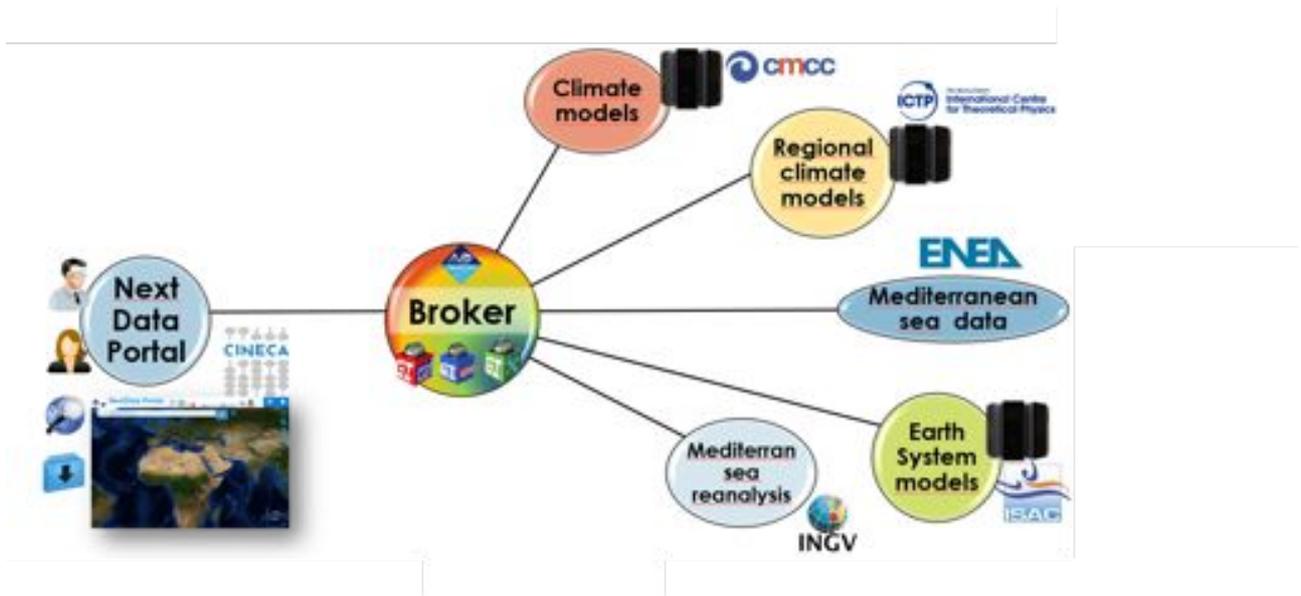


Figure 1 NextData brokering architecture, showing the main components depicted inside balloons (ND Portal, ND Broker and ND Data Providers) and the interconnections between them drafted using black lines.

The Broker acts as a mediator between the NextData portal and the available sources. Discovery and access functionalities are indeed mediated by the Broker across all the data sources, upon a client request.

The NextData Portal makes use of a simple JavaScript library to interface with the broker. The library is called GI-API and the current version is v1.2.9. The following sections describe the latest functionalities of the NextData Portal, implemented by GI-API and ultimately by the NextData Broker. The list of all the implemented functionalities of the NextData Portal is depicted in Annex A.

2. Discovery functionalities

The discovery functionalities allow users to find the NextData resources of interest. The following sections describe the main functionalities: data sources browsing, query based search, results filtering and semantic search.

2.1 Data Sources Browsing

NextData Providers content can be hierarchically browsed through the Data Sources window, as shown in Figure 2. All the sources are listed, for each one both a title and a graphical icon representing the protocol used by the Provider to publish the data are displayed. Users can expand each node list to explore the content of the Data Provider of interest. The hierarchical view shown by this component reflects the organization of contents implemented by each Data Provider. Folders and subfolders can be navigated until leaf datasets are reached: they are represented by a file icon; they can be further evaluated by showing their metadata and possibly accessed.



Figure 2 Data Sources Browsing using a tree component.

2.2 Query Based Search

Users can issue queries using the NextData Portal by specifying a set of constraints that can be a combination of:

- **Search Term.** The Broker matches resources having the free text provided as input by the users in one or more of these metadata fields: Title, Keywords, Description.
- **Spatial-Temporal Extent.** The broker matches resources that have their spatial-temporal domains (e.g. bounding box and temporal extent) overlapping or contained in the spatial-temporal extent chosen by the user.
- **Data Sources Subset.** The broker can constraint the search on a given set of Data Providers, selected by the user by checking them in the Data Source panel.

Results are shown as a paginated list (see Figure 3). For each entry, some basic metadata such as resource title, geographic extent and description is provided. By clicking on the information button, extended metadata for each result can be further viewed and evaluated.

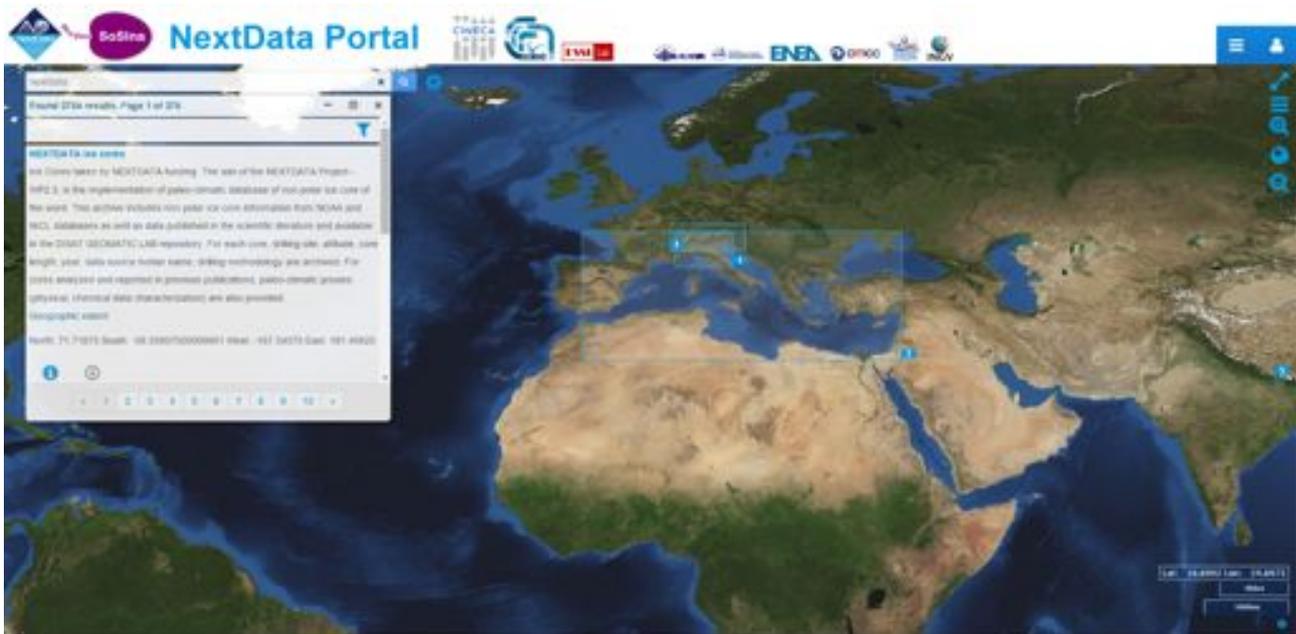


Figure 3 Query Based Search. Results are shown on the map in an aggregated form.

2.3 Results Filtering

After a search is performed, it is possible to prune the possibly long list of results using the filtering functionality by clicking on the funnel icon on the top of the list. Aggregation is performed on a set of metadata elements: keywords, formats, source and protocol. The aggregated results panel is divided into four sections, one for each metadata element, and in each section an aggregated list of the results (by the specific element) is shown.

Example given (see Figure 4), the keyword section lists all the keywords that are present in the matched results. Each keyword (e.g. “salinity”) is accompanied by the number of results which have that specific keyword in their metadata (e.g. “328”). By clicking on a particular keyword, the actual list of these resources is shown, enabling subsequent refinements or evaluation and access.

2.4 Semantic Search

A semantic search can also be performed, by using the free text field and specifying the type of semantic relation in the Semantic Options panel. Three types of relations are available: *specific*, *general*, or *similar*. To execute this type of queries the Broker makes use of a third party service which exposes a set of aligned controlled vocabularies. The Broker expands the single term inputted by the user making use of the semantic service, in order to create a set of new queries, having terms that are semantically related to the original one.

Different sets of results are shown in different tabs (see Figure 5), one for each expanded term.

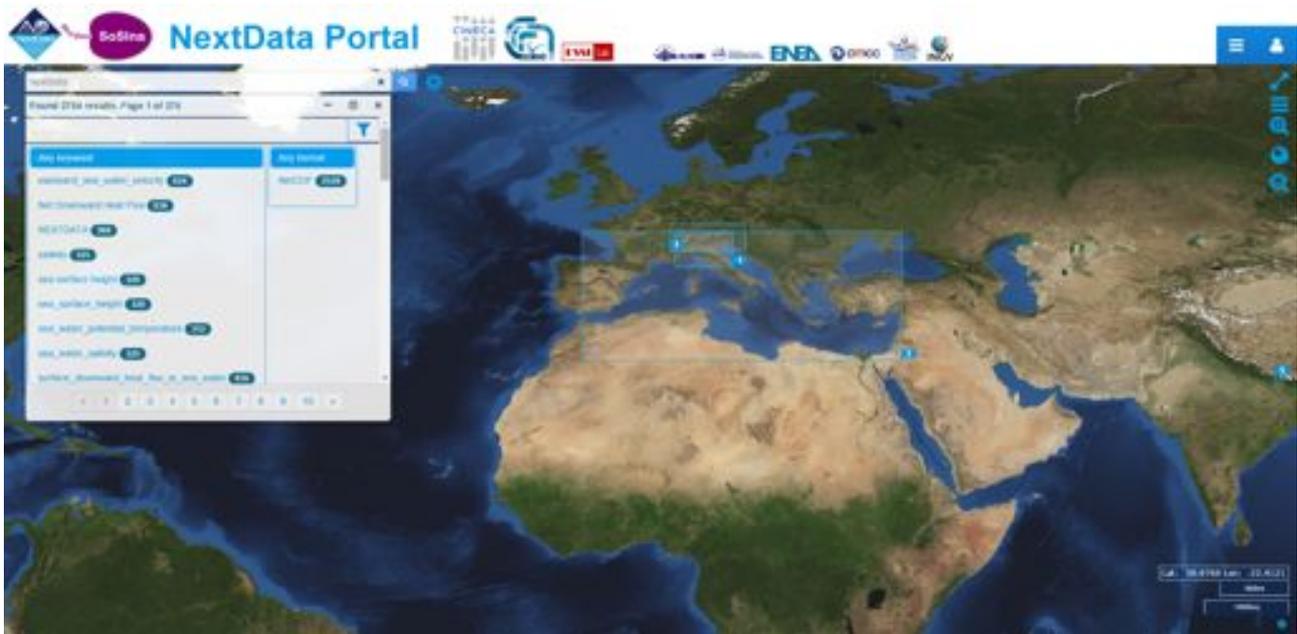


Figure 4 Results Filtering. An aggregated results view can be used to further refine the results list, by adding filters.

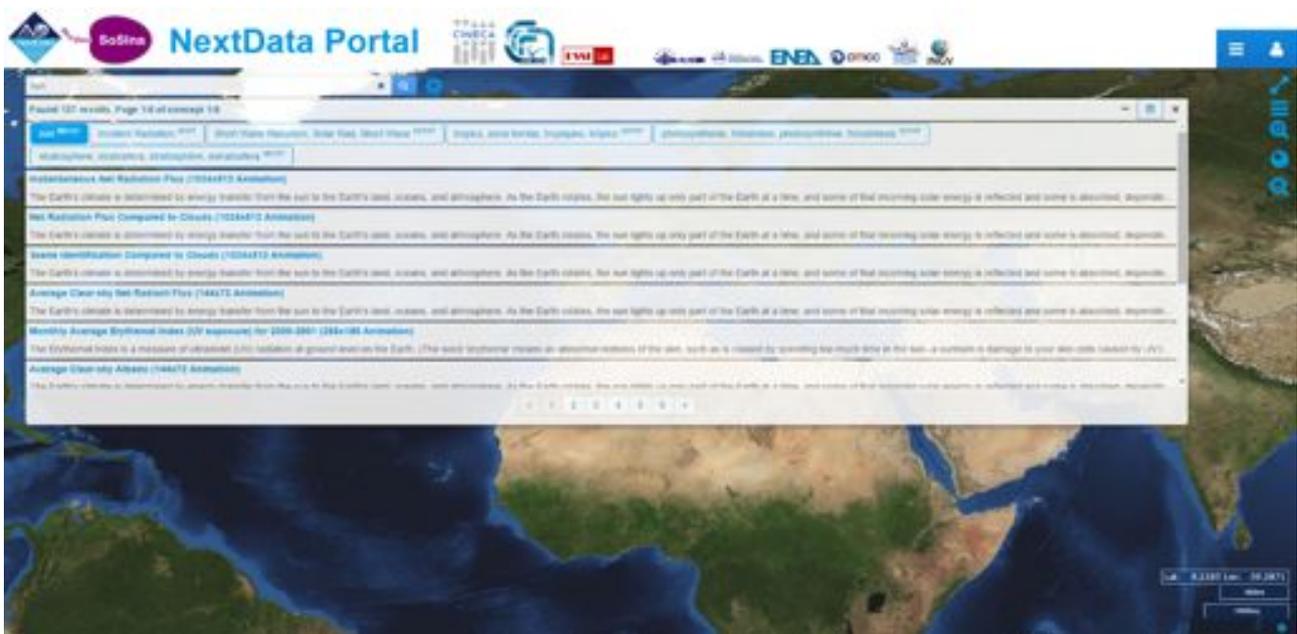


Figure 5 Semantic Search: a single search term is semantically expanded, giving rise to a set of semantically related queries and result sets.

3. Access functionalities

One or more resources of interest for the users represent the output of the discovery phase. The access phase consists in accessing (i.e. downloading) the actual data. The pre-requisite is, of course, data availability and the public endpoint is correctly described in the metadata distribution information.

3.1 Direct download

The Broker can perform direct download of a selected resource (users have to click on the direct download button for the desired resource). In this case data is download in the Common Grid Environment (CGE) - a combination of format, resolution, Coordinate Reference System, extents - provided by the source Data Provider.

3.2 Mediated access

The Broker can also perform mediated access: user can request a specific resource in a desired CGE. It will be up to the broker to return the resource accordingly formatted. In case the Data Provider doesn't support access to a resource in the user requested CGE, then the Broker implements the missing transformations, by supplementing the source access functionalities. Example Given, a resource published on a Web Coverage Service may be originally available for download in a raster format only, such as NetCDF. User may instead request a pictorial format, such as an image (e.g., png). In this case the broker will take care of the needed format transformation.

Figure 6, Figure 7 and Figure 8 demonstrate mediated data access using the Portal. In Figure 6 search results are shown. For some of them data preview is also available, in order to give a quick overview to the User. A user can proceed to the download clicking on the download button: the access parameter panel is then shown (Figure 7). Here spatial and temporal extent, format, resolution and CRS can be chosen. Finally the dataset is downloaded in the desired CGE and user can further exploit it with his favorite application (Figure 8).

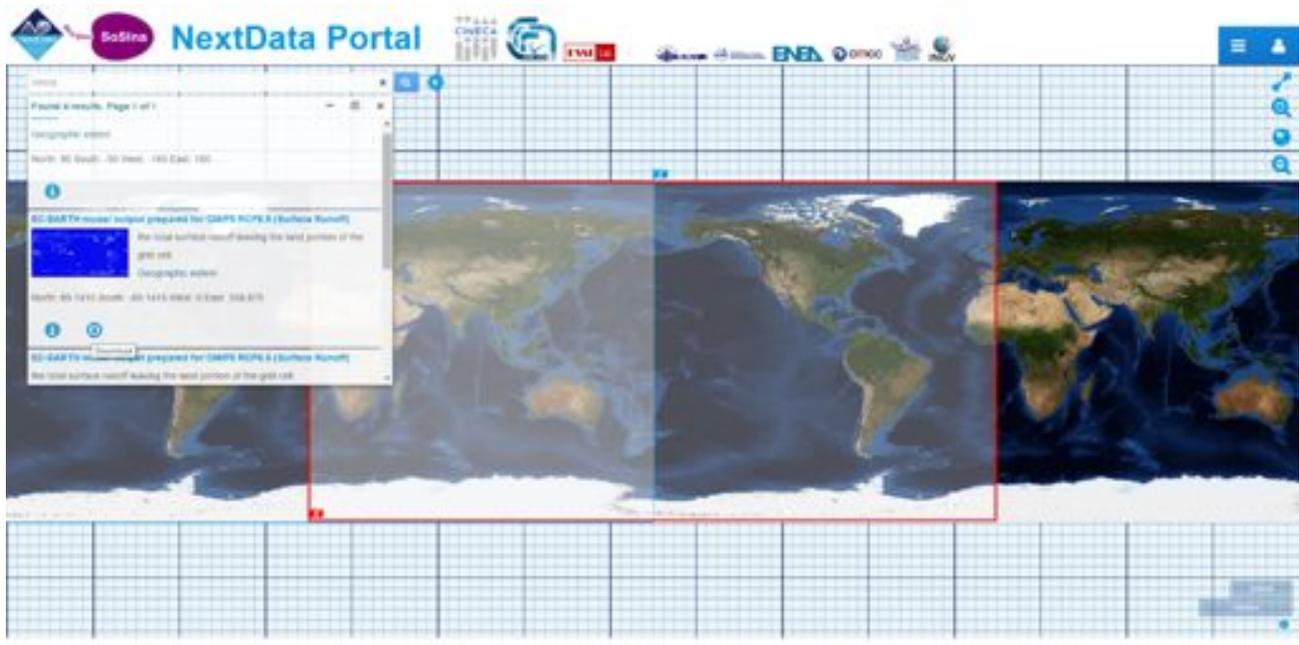


Figure 6: Search results showing data previews.

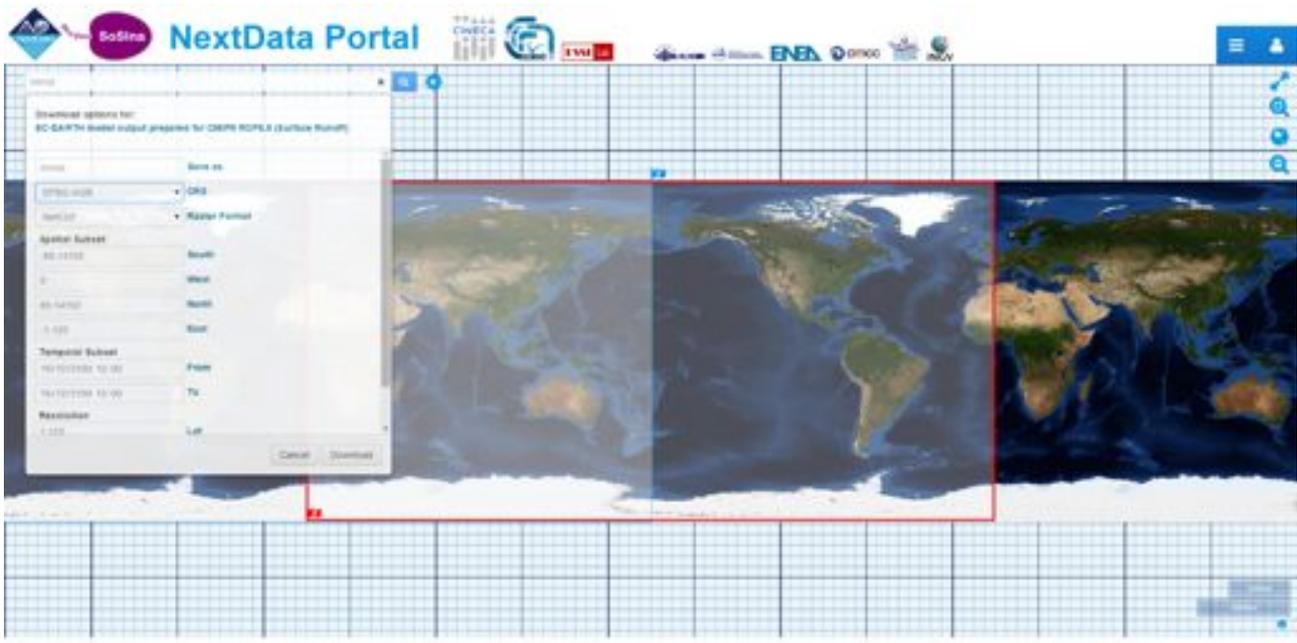


Figure 7 Mediated access. User can specify the desired access parameters to download the datasets according to a particular Common Grid Environment (i.e. a combination of Format, Resolution, Extent and Coordinate Reference System).

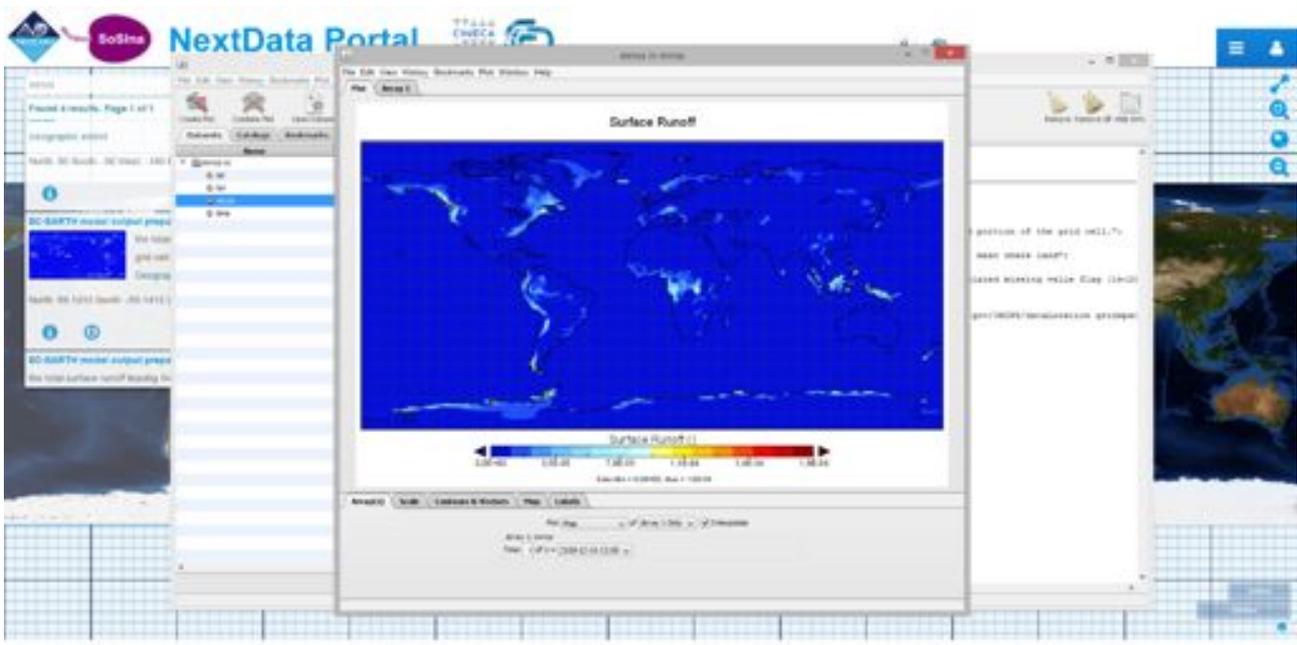


Figure 8 Data access result. User is able to view and further exploit the just accessed NextData dataset.

3.3 Data Policy

According to the Data Management Plan (including the Data Policy), the NextData Portal does not store any data but broker contributing data systems via its brokering middleware. Therefore, only metadata are centrally harvested and harmonized.

The NextData harmonized metadata are provided as Open Data to any “client” system that can access them via public APIs implemented by the NextData brokering middleware.

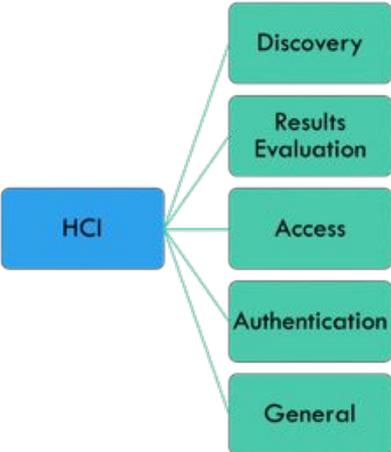
The NextData data policy is defined at the level of the general governance of the system of systems negotiating that with each of the contributing (and brokered) data systems –see Figure 1.

Annex A. Portal Functionality schematic

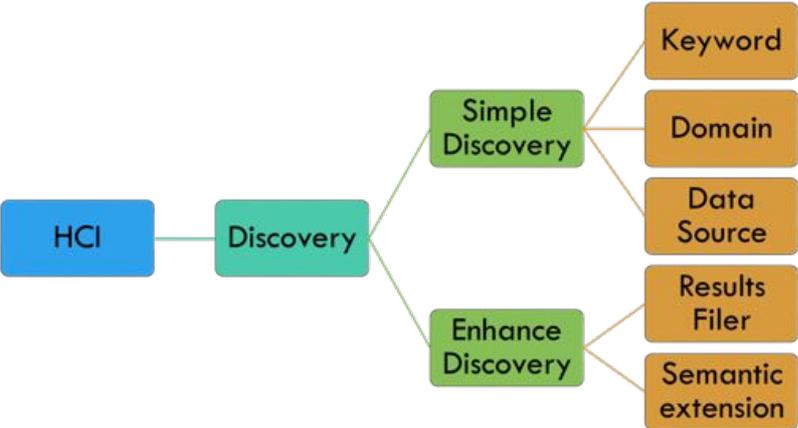
A1: Human Computer Interaction (HCI) Overview



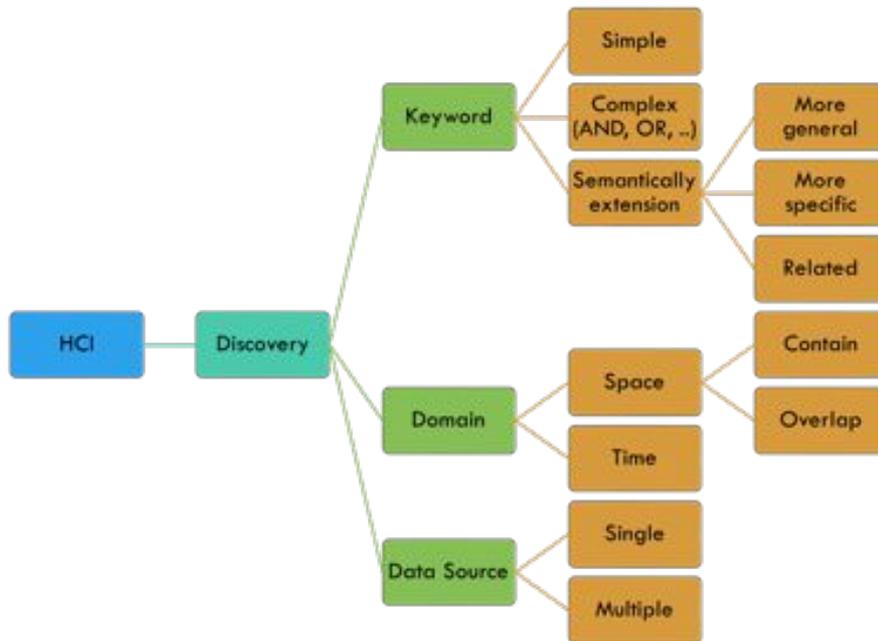
A2: HCI first level



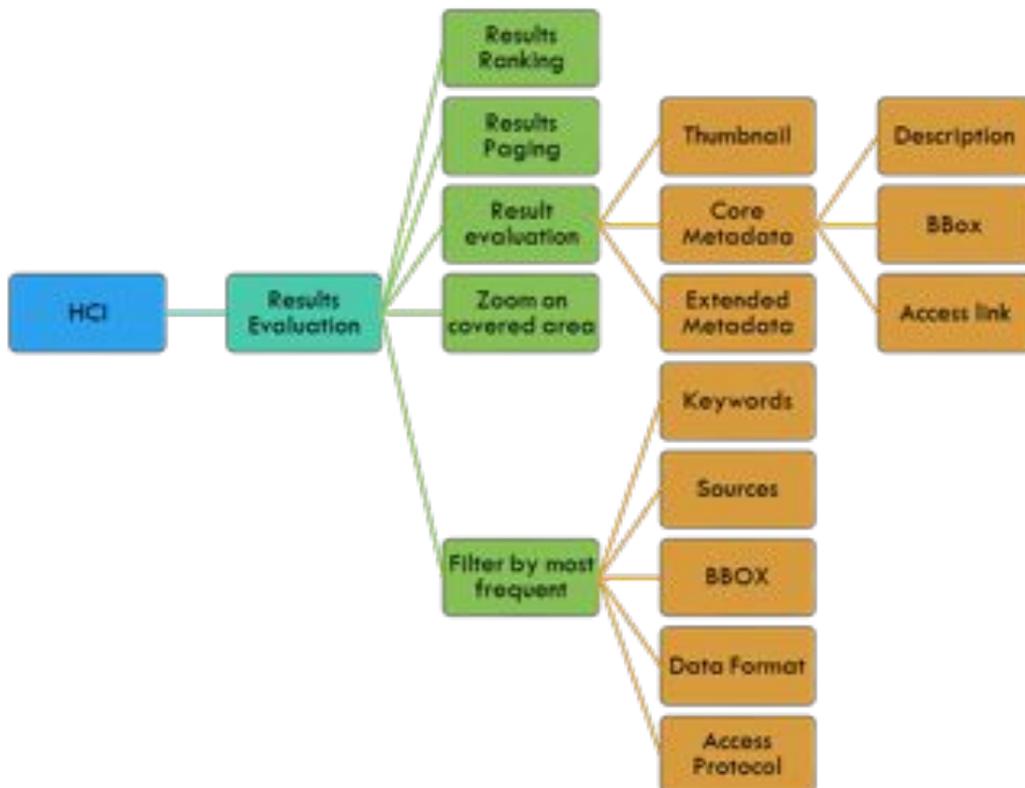
A3: Discovery high-lev functionalities



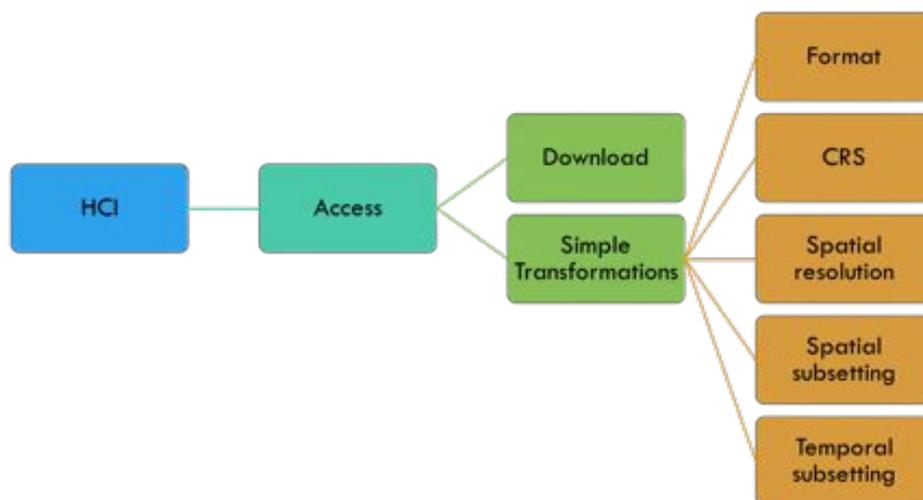
A4: Simple Discovery high-lev functionalities



A5: Discovery Results Evaluation functionalities



A6: HCI Access and simple Transform



A7: HCI General Functionalities

