Project of Strategic Interest NEXTDATA

Deliverable D2.7.7.a
Extended Brokering Middleware
(ND-SoS-Ina)

WP Coordinator: Stefano Nativi

CNR – IIA

Document identifier:	SoSIna D2.7.7.a
Date:	April 2015
Editors:	Enrico Boldrini, Fabrizio Papeschi, Stefano Nativi
Activity:	SosIna WP3
Lead Partner:	CNR
Document status:	Final
Document link:	http://essi-lab.eu/nextdata/sosina/?page_id=68
Diffusion	Internal
Abstract	This document presents the extended brokering framework that enables the interoperability and harmonization functionalities of the NextData General Portal.
Keywords	Brokering, Interoperability, Middleware, Data Access, Data Management

History

Document version	Date	Authors
1.0	Apr 2015	Enrico Boldrini –CNR-IIA Fabrizio Papeschi CNR-IIA Stefano Nativi - CNR-IIA

Abstract

This document presents the extended brokering framework that enables the interoperability and harmonization functionalities of the NextData General Portal.

Keywords

Brokering, Interoperability, Middleware, Data Access, Data Management.

TABLE OF CONTENTS

1 Brokering architecture

The diagram in **Fig**. *1* shows the current deployment of the NextData architecture. The principal components are:

- NextData 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.
- NextData Broker: It is the middleware which connects the Portal with the available data sources. It acts as a mediator of discovery and access functionalities. 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 normalized by the Broker and sent back to the Portal, to be displayed to the users in a convenient way.
- NextData 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)
  - URT EVK2CNR (GeoNetwork, Web Map Service, includes some Special Projects)
  - UNIMIB (Italy-2k data, ice cores)

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.
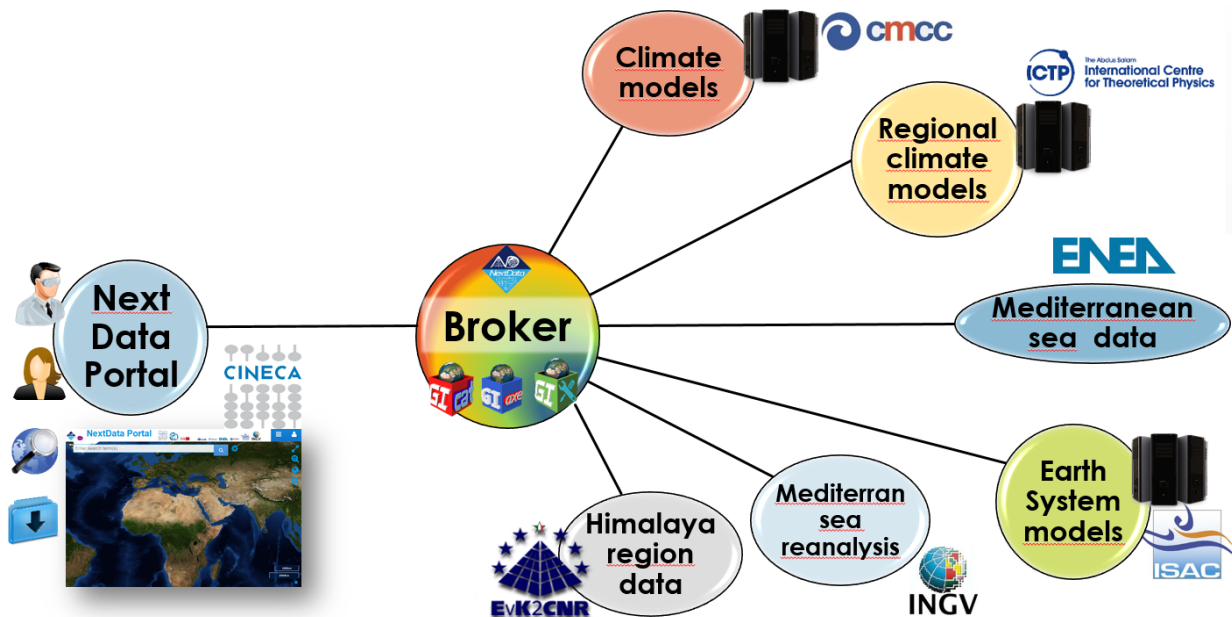

Fig. 1 NextData brokering architecture, showing the main components depicted inside balloons (ND Portal, ND Broker and ND Data Providers) and the interconnections between them drafted using black lines.

The NextData General Portal makes use of a simple JavaScript library to interface with the Broker. The library is called GI-API and is currently at version 1.2.9. The following sections describe the latest functionalities of the NextData Portal, implemented by GI-API and ultimately by the NextData Broker.

1.1 Discovery functionalities

User can find NextData resources of interest making use of the discovery functionalities. The following sections describe the main ones: data sources browsing, query based search, results filtering and semantic search.

1.1.1 Data Sources Browsing

NextData provider content can be hierarchically browsed through the Data Sources window, as shown in **Fig. 2**. All sources are listed; for each one of them a title is shown and a graphical icon represents the protocol used by the provider to publish the data. 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 content 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.


Fig. 2. Data Sources Browsing using a tree component.

1.1.2 Query Based Search

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

- **Search Terms.** The Broker matches resources having the free text inputted by the users in one or more of these metadata fields: Title, Keywords and Description.
- **Spatial-Temporal Extents.** The broker matches resources that have their spatial-temporal domains (e.g. bounding box and temporal extent) overlapping or contained in the spatial-temporal extents 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 **Fig. 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 results can be further viewed and evaluated.
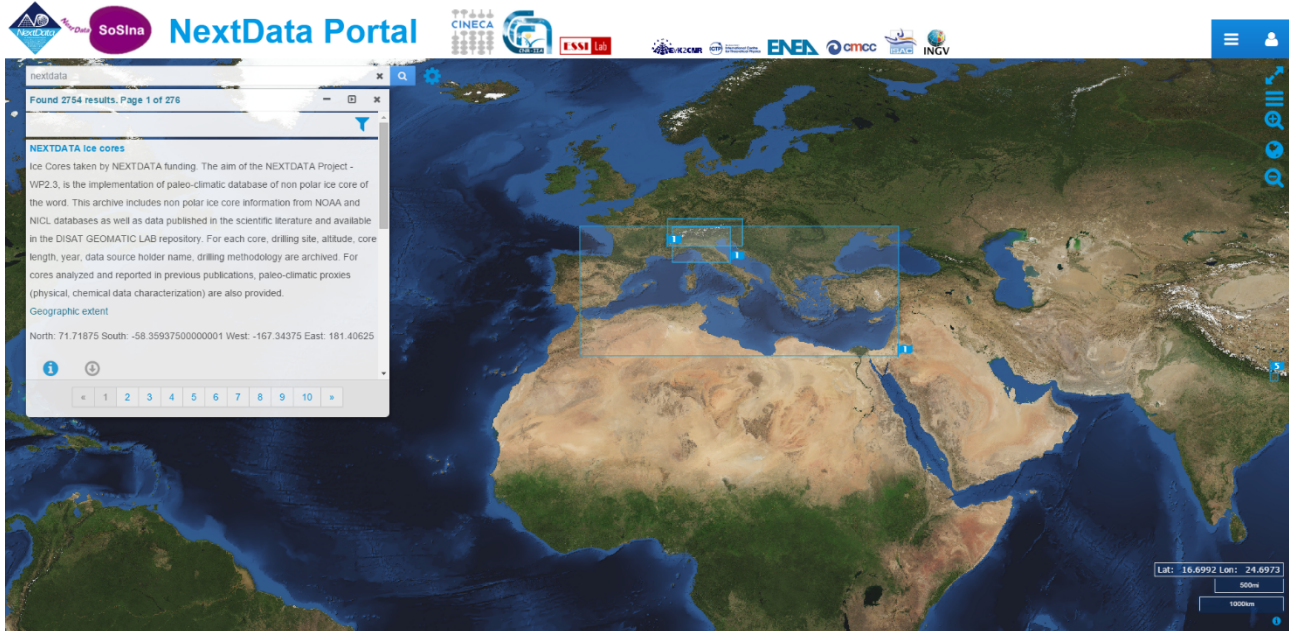

Fig. 3. Query Based Search. Results are shown on the map in an aggregated form.

1.1.3 Results Filtering

After a search is performed, it is possible to prune the (often long) list of results using the filtering functionality by clicking on the funnel icon on top of the list. Aggregation is performed on a set of metadata elements: keywords, formats, source and protocol. The aggregated results panel is divided in 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 **Fig. 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 particular keyword in their metadata (e.g. "328"). By clicking on particular keywords, the actual list of these resources is shown, enabling subsequent refinements or evaluation and access.
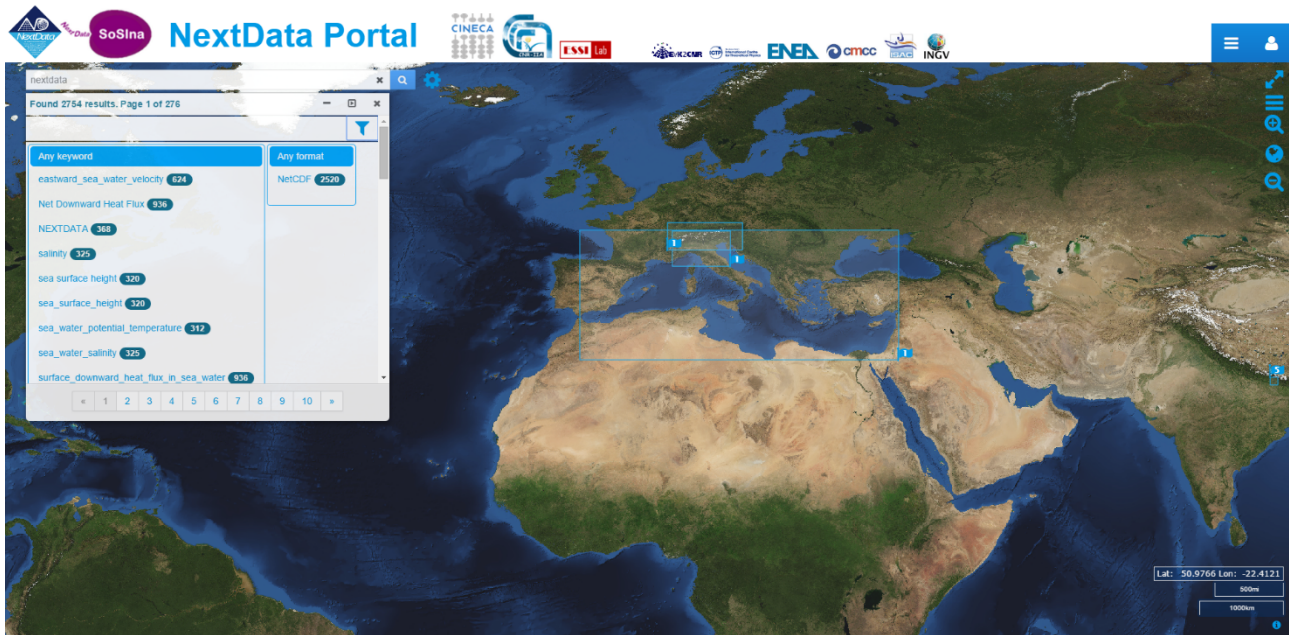
Fig. 4. Results of Filtering. An aggregated view can be used to further refine the result list, by adding filters.

1.1.4 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: more specific, more general, 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 result sets are shown in different tabs (see **Fig. 5**), one for each expanded term.

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

1.2 Access functionalities

One or more resources that are of user interest 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 that data is available and the public endpoint is correctly described in the metadata distribution information.

1.2.1 Direct download

The Broker can perform direct download of a selected resource (users have to click on the direct download button shown for the desired resource). In this case data is download in the Common Grid Environment (a combination of format, resolution, CRS, extents) provided by the source data provider.
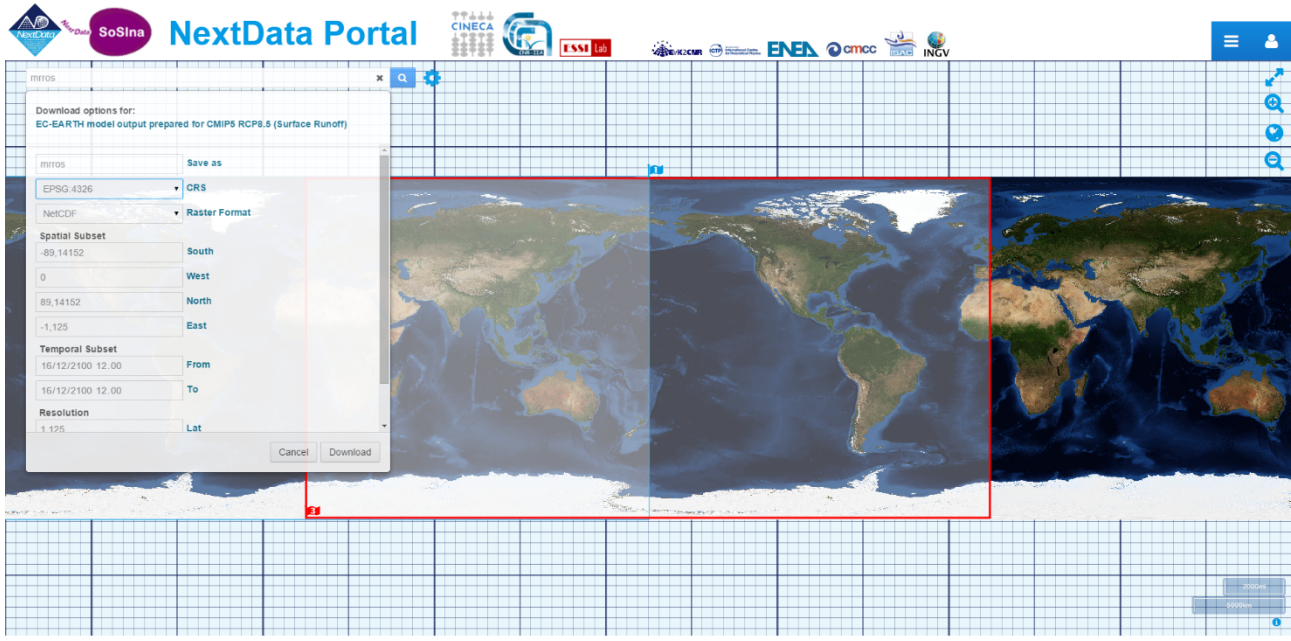
1.2.2 Mediated access

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 of 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. Users may instead request a pictorial format, such as image/PNG. In this case the Broker will take care of the needed format transformation.

Fig. *6*, **Fig. *7*** and **Fig. *8*** demonstrate mediated data access using the Portal. In **Fig. *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 (**Fig. *7***). Here spatial 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 (**Fig. *8***).
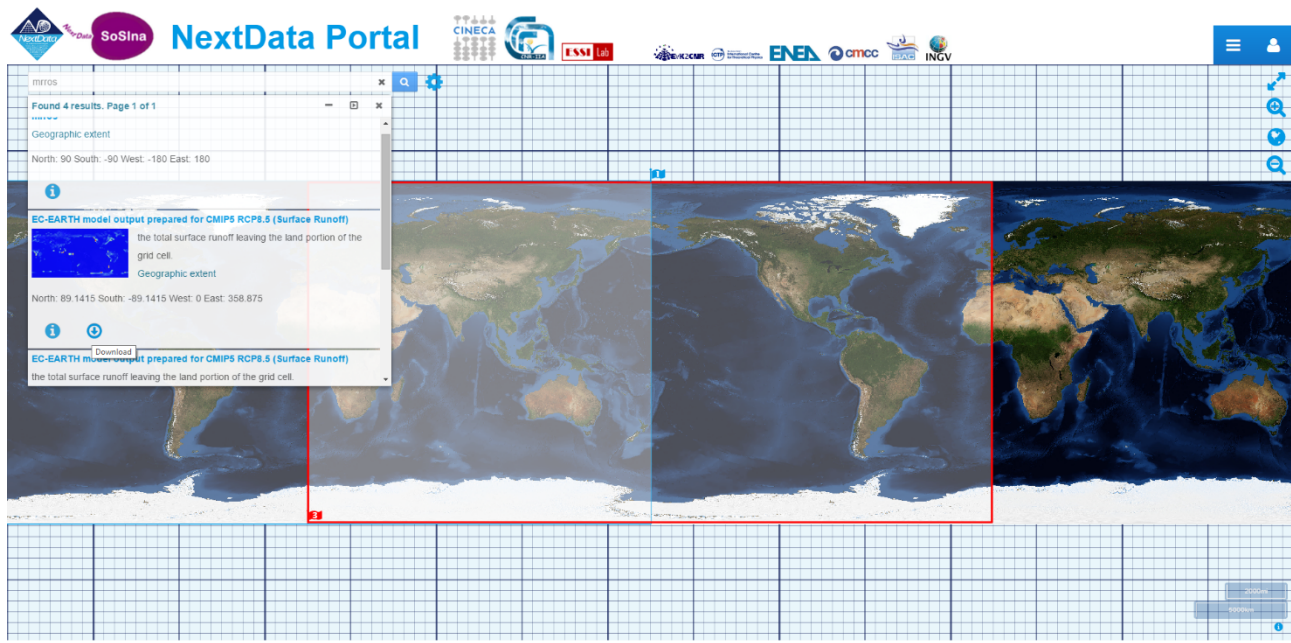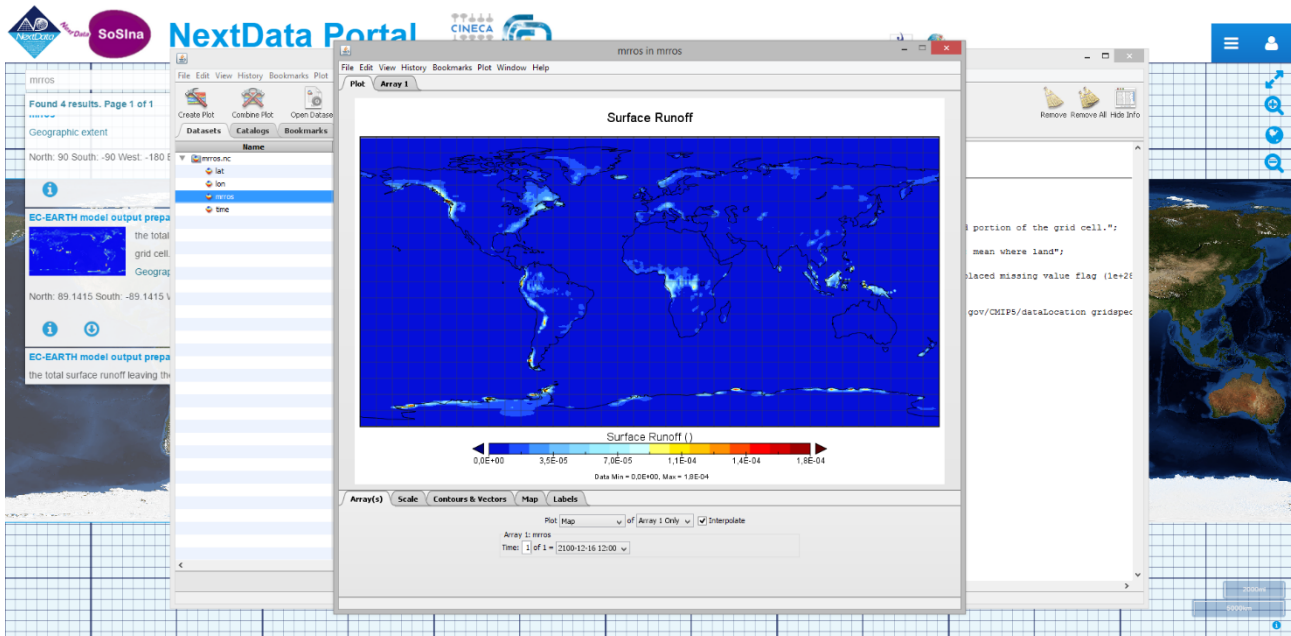

Fig. 6. Search results showing data previews.

Fig. 7. Mediated access. Users 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).

Fig. 8. Data access result. Users are able to view and further exploit the just accessed NextData dataset.