



Requirements and System Architecture (ND-SoS-Infra)

DELIVERABLE D2.1 (CNR-IIA, CINECA) - General Requirements
 DELIVERABLE D2.2 (CNR-IIA, CINECA) - General System and Interoperability
 Architecture

Document identifier:	D2.2-3
Date:	Dic 2013
Editors:	Stefano Nativi,
Activity:	WP2 - Task 2.1 - 2.2 - 2.4
Lead Partner:	CNR, CINECA
Document status:	draft
Document link:	
Diffusion	Internal

History

Document version	Date	Authors
1.0	Dic 2013	Stefano Nativi - CNR-IIA Paolo Mazzetti - CNR-IIA Marco Rorro - CINECA

Abstract

This document presents the **NextData SoSIna System requirements and Interoperability Architecture** providing the necessary inputs for the Web portal design and development.

Keywords

System Architecture, Requirements, Brokering.

TABLE OF CONTENTS

- [1. Introduction: the NextData Users' needs](#)
- [2. General System and Interoperability Architecture](#)
[Leading Principles and system architecture](#)
- [3. Next Data Multidisciplinary Requirements](#)
[Functional Requirements](#)
[Non-functional Requirements](#)
- [4. Use Cases](#)
- [5. Future extensions](#)
- [6. Conclusions](#)
- [7. References](#)

1. Introduction: the NextData Users' needs

Having interacted with NextData Scientists and with many stakeholders working on Global Changes and the Earth System Science, as well as considering the requirements coming from other relevant projects and initiatives (like GEOSS, INSPIRE, Copernicus, Belmont Forum, Earth Cube, and some FP7 and ESA funded programmes -including EuroGEOSS, GEOWOW, Med-SUV, MEDINA, Prod-Trees, GeoViQua, RECODE, EO-POWER, EU-BON, COOPEUS, ODIP, SeaDataNet II, etc.) a first set of common requirements includes the following needs for the NextData e-infrastructure:

- To achieve an initial operating capability, including the development of the services that are necessary to make it possible to discover, view, access and use the information resources made available by the project partners in the different thematic areas.
- To undertake the research necessary to further develop thematic partner infrastructures to provide access not just to data but also to selected analytical models and workflows.
- To discover and access to resources made available through heterogeneous and distributed systems and DBs.
- To rank and classify discovered results coming from different distributed catalogs and systems. Resources could be ranked either by considering: (i) the publication as GEOSS Data CORE, (ii) the quality of the metadata (presence of title, description, URI, resource accessibility, availability of resource etc), and (iii) the resource fit a particular purpose/scenario.
- To address the semantic heterogeneity among the terminology used by different scientific communities.
- To access data according to the needed spatial and temporal resolution, format encoding and Coordinate Reference System (from any of the available download services) using the user's preferred service interface.
- To access and harmonize data through controlled data schema mappings across different provider repositories.
- To implement a scalable and flexible infrastructure that seamlessly allows users to compose available processes to build complex business processes and execute them.
- To adopt a common methodology to describe processes.
- To execute either single or chained/compound processes through a Process Orchestration Engine.
- To perform quality assessment of the resources discovered in the Next Data infrastructure.
- To include provenance information and assure science reproducibility.
- To improve the Open Data availability, supporting easy registration, discovery and access of Open Data resources.
- To address access condition, through registration and authorization functionalities
- To facilitate public participation and interaction with citizen-provided content.

- To create and implement an easy to use and flexible Web Portal.
- To address non-functional requirements such as scalability, performances and different security levels.

From these requirements, a set of functional and non-functional requirements have been recognized to design the NextData Web Portal and the NextData infrastructure architecture, more generally.

2. General System and Interoperability Architecture

It is possible to introduce a first list of high-level “system and software requirements”, building on the Users’ needs as well as on the interoperability requirements for interconnecting the NextData infrastructure to the significant international initiatives and programmes (GEOSS, INSPIRE, Copernicus, Belmont Forum, Earth Cube, EoE, Unep-live, WIS, etc.). In particular, NextData aims to provide a significant Italian contribution to the GEOSS Ecosystems task by also interconnecting its e-infrastructure to the GEOSS Common Infrastructure (GCI) [R1]. Therefore, NextData is required to design, implement, and operate a multi-disciplinary e-infrastructure, applying the interoperability principles recognized by GEO.

Leading Principles and system architecture

The NextData infrastructure is called upon to apply the following principles, characterizing the System of Systems (SoS) approach successfully introduced with the GCI [R2]:

- A. Build on existing system/network infrastructures.
- B. Supplement but not supplant system/network (i.e. Community) mandates and governance arrangements.
- C. Address heterogeneity to lower entry barriers.
- D. Be flexible and extensible to:
 - i. Interconnect new system/network infrastructures.
 - ii. Sustain and advance the achieved interoperability.
 - iii. Allow each system/network (i.e. Community) infrastructure to evolve.

Therefore, the first set of high-level system and software requirements includes the need to build on the existing (or next coming) information systems published by the NextData “Resource Providers”.

To be interoperable, each NextData “Resource Provider” needs to publish one or more Internet based services/interfaces to make it possible to discover and access their shared resources. Virtually, there is no constraint on the types of published services/interfaces: semantically speaking, they can be either very simple (e.g. FTP, Web Accessible Folder, DB Web interface)

or more complex (e.g. OGC/ISO Web services, Community SOAP-based/REST services, linked-data services).

On the top of these “Community” information systems, the NextData e-infrastructure will be realized. That will be in charge of three main tasks:

- A. The interconnection of the heterogeneous information systems provided by the NextData Partners, in a transparent way.
- B. The harmonization of the data and metadata model diversities;
- C. The implementation of a set of multidisciplinary discovery, evaluation and access services -including the publication of a general Web Portal.

Figure 1. depicts the described architectural approach.

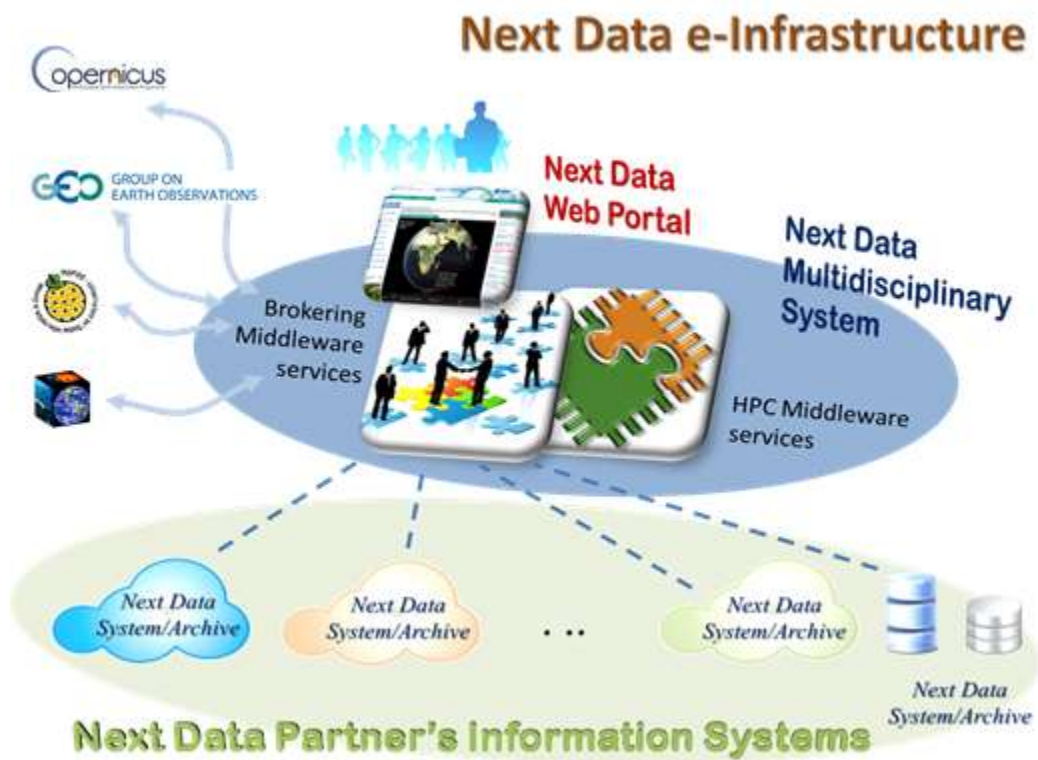


Figure 1. NextData e-infrastructure architecture

3. NextData Multidisciplinary Requirements

Functional Requirements

The NextData e-infrastructure shall implement a set of multidisciplinary functional requirements including:

- (a) Architectural Interoperability
 - access to heterogeneous and distributed systems (supporting service-oriented, resource-oriented and event-based architectures);
- (b) Discoverability and Evaluation
 - heterogeneous metadata access and harmonization to allow evaluation;
 - classification and ranking of discovered resources;
 - quality and provenance information support (including users' feedbacks)
- (c) Accessibility
 - Heterogeneous data access and harmonization (where requested);
 - Transformation services to supplements data providers capabilities;
- (d) Semantics Interoperability
 - Heterogeneous semantic assets access and use to support semantic discovery;
 - Semantic assets alignment (where requested);
- (e) Processing
 - possibility to compose processes and scientific models;
 - community workspaces integration;
- (f) Organization Interoperability
 - security and trustability services;

Non-functional Requirements

The NextData e-infrastructure shall implement a set of non-functional requirements (quantitative criteria must be decided in the due course) as well, dealing with the following main challenges [R3]:

- (a) User Interface and Human Factors
 - What type of user will be using the system?
 - What sort of training will be required for each type of user?
- (b) Documentation
 - What kind of documentation is required?
 - What audience is to be addressed by each document?
- (c) Performance Characteristics
 - Are there any speed, throughput, or response time constraints on the system?
 - Are there size or capacity constraints on the data to be processed by the system?
- (d) Quality Issues
 - What are the requirements for reliability?
 - What is the acceptable system downtime per 24-hour period?
- (e) System Modifications

- What parts of the system are likely candidates for later modification?
 - What sorts of modifications are expected?
- (f) Resources and Management Issues
- How often will the system be backed up?
 - Who will be responsible for system maintenance?

Data Sharing requirements

In keeping with the Data policy of NextData [4], the multi-disciplinary e-infrastructure shall apply the following sharing principles:

- There will be full and open exchange of data, metadata and products shared within NextData, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

4. Use Cases

Several use cases are possible with the NextData e-infrastructure; however, the first and simplest one to be supported is the following (see Figure 2.):

1. NextData User generates and issues a discovery request (i.e. query)
 - a. The NextData e-infrastructure answers with a set of matching results coming from all the brokered resources.
2. NextData User browses and evaluates the result set, asking the system to visualize dataset information (i.e. metadata) and a portrayal.
3. NextData User requires to access one or more evaluated datasets, specifying a common spatial and temporal domain and data format (i.e. common grid environment) -to be improve their usability.
 - a. The NextData e-infrastructure finalizes all the transformations required to return the selected datasets according to the specified common grid environment.

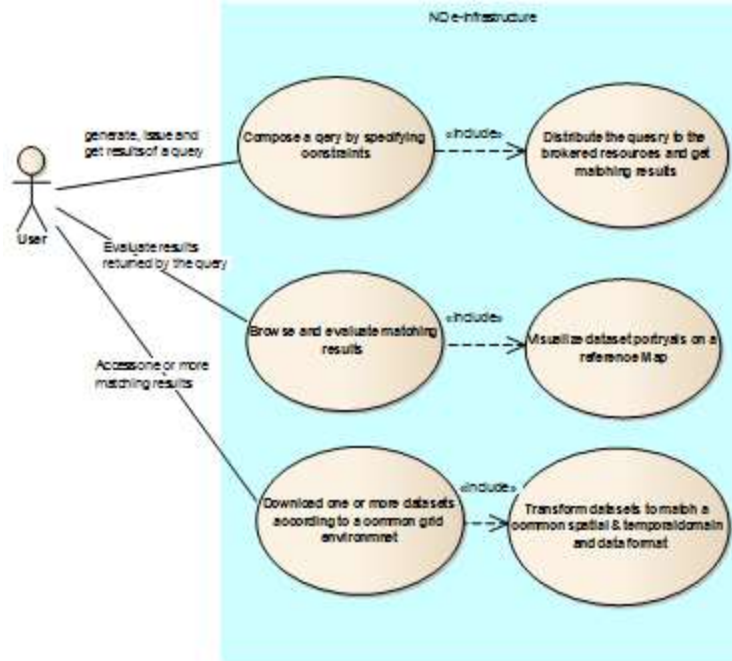


Figure 2. The NextData “discovery & access” basic Use Case

5. Future extensions

In an AGILE software development approach, requirements are often refined and new ones are introduced during all the project life-time. Therefore, this document will be updated after the first major development cycle.

6. Conclusions

In this document we have presented the high-level system and software requirements for the architecture of the NextData e-infrastructure.

Specific functional and non-functional needs are listed and briefly discussed.

7. References

R1	GEOS Web Site: http://www.earthobservations.org/geos.shtml
R2	Brokering Approach: S. Nativi et al., IEEE JSTARS manuscript 2013.

R3	Non-functional requirements: CSEE, University of Maryland
R4	Policy Data of NextData: