Project of Interest
"NextData"

Research project :

NextData System of Systems Infrastructure (ND-SoS-Ina)

Coordination: Stefano Nativi

Istituto sull'Inquinamento Atmosferico (IIA)

Work Pachage :
WP 2.6

TITLE OF THE PROPOSED PROJECT:
NextData System of Systems Infrastructure (ND-SoS-Ina)

Project duration: 2013-2015 (28 months)
Start date: Jun 2013
End date: Sep 2015

Scientific coordinator of the proposed project:
Stefano Nativi (CNR-IIA)

CNR Institute coordinating the proposed project:
Istituto sull'Inquinamento Atmosferico (IIA)

Participating units:
Unit 1 (CNR coordinating Institute): IIA (scientific responsible: Paolo Mazzetti)
ESSI-lab of CNR-IIA provides the following expertise: multidisciplinary interoperability, cyber-science, earth and space science informatics.
Unit 2: CINECA (scientific responsible: Marco Rorro)
Middleware for HPC group of CINECA complements the ESSI-lab expertise by providing: experience in the developing HPC middleware components, hosting infrastructure, and operational services.

1. GENERAL INFORMATION

Abstract of the proposed project

ND-SoS-Ina is a research and development project in keeping with the activities and goals of the NextData project. This contributes to reinforce the Italian and CNR presence in the innovative and important (see Horizon 2020) research areas of cyber-science and multidisciplinary interoperability. The NextData cyber(e)-infrastructure adopts a System of Systems approach based on a brokering architecture. The project is comprised of three distinct phases and related milestones: (a) the first prototype of the NextData System of Systems infrastructure, implementing the core functionalities; (b) the consolidated version of the NextData System of Systems infrastructure, implementing advanced functionalities; (c) the final and operative NextData System of Systems infrastructure for data and information sharing and publication.

Main goals of the project

ND-SoS-Ina main goal is to design, develop, and operate in production the NextData cyber(e)-infrastructure for the publication and sharing of its resources (i.e. data, services, controlled vocabularies, models, etc.). ND-SoS-Ina cyber(e)-infrastructure realizes the NextData general Portal implementing the interoperability among the data archives carried out by NextData. This is pursued by applying the brokering architecture and developing framework first introduced by the EC-FP7 EuroGEOSS project and then consolidated by the international programme GEOSS (Global Earth Observation System of Systems) of GEO (Group oh Earth Observation).

Expected results of the project

ND-SoS-Ina project will deliver the NextData cyber(e)-infrastructure adopting a three phases approach: (a) the delivery of first infrastructure prototype providing the core functionalities (i.e. discovery, evaluation, access, and visualization); (b) the delivery of the consolidated version of the infrastructure, implementing advanced functionalities (e.g. advanced scalability and robustness, preservation); (c) the final and operative infrastructure. In particular, the following infrastructure components will be provided (and/or integrated) by the project: (a) the NextData General Portal providing the Human-Computer Interaction functionalities for all the resource sharing services; (b) the Broking Framework for implementing flexible and low entry level interoperability among different disciplines; (c) the High Throughput Framework for implementing middleware data services; (d) middleware components to manage geospatial data and metadata in a user-friendly way. ND-SoS-Ina infrastructure services will enable the data sharing among the NextData systems and archives in a

transparent, flexible, and sustainable way. Finally, multidisciplinary applications will be able to use the NextData cybe(e)-infrastructure as a low entry barrier access-point for other global infrastructure (e.g. GEOSS, INSPIRE, WIS, PANGAEA, GBIF, etc.).

Role of the different units

CNR-IIA

CNR-IIA coordinates the entire project. Besides the Work Package on "Coordination and Management" (WP1), CNR-IIA coordinates WP2: "Requirements and System Architecture Design" and (WP5) "Dissemination". Finally, CNR-IIA leads Task 3.1: "System of Systems middleware extension and integration" and the "General operation of the Web Portal" activity of Task 4.2.

CNR-IIA provides the Brokering Framework, designed and developed first for EuroGEOSS and then for GEOSS, leading all the activities aiming to its extension for and integration with the other components (in particular the General Portal) of the NextData infrastructure. CNR-IIA leads the NextData interoperability infrastructure design, in keeping with the principles and technologies adopted by relevant global and multidisciplinary programmes, like: GEOSS and EarthCube. CNR-IIA will particularly disseminate the ND-SoS-Ina project results in the GEOSS Community.

Finally, CNR-IIA will coordinate with the NextData leader and the other co-Principal Investigators to assure that:

(a) the General Portal will be suitable for the project purposes;
(b) user requirements effectively identify the Users' needs for climate data management, processing and visualization.

CINECA

CINECA (Consorzio Interuniversitario) is the largest Italian supercomputing centre equipped with the most advanced hardware resources. CINECA's mission is to enable, while facilitating, scientific researchers and private customers exploit HPC systems in a profitable way and to understand and anticipate future user needs. The "Middleware for HPC Services" group, as part of the SuperComputing Applications and Innovation department, will be directly involved into the project bringing its experience and expertise in developing complex middleware systems. Major contribution will be devoted to the implementation of the Web Portal, the deployment and hosting of the final service, and the enhancement of overall system performance. CINECA will coordinate the development of the system taking the leadership of the WP3 "System Development, Integration, and Test" and the operation of the final service on its infrastructure through the WP4 "Operating the System".

2. DETAILED PROJECT DESCRIPTION

State of the art and motivations

Global sustainability research requires an integrated multi-disciplinary effort underpinned by a cyber(e)-infrastructure able to harness big data and heterogeneous information systems across disciplines (Nativi and Fox, 2010). Two approaches are possible to achieve the interoperability required across such systems and data: federating, and brokering. The former proved to be appropriate to single discipline or "controlled" domain environments, but has several shortcomings for "complex" (i.e. global, multidisciplinary, multi-organizational, quickly evolving, etc.) domain environments (Nativi et al., 2012) (Nativi et al., 2013). On the contrary, the brokering approach has showed to be more scalable and effective to realize global System of Systems and Network of Networks. The brokering approach and related architecture were introduced first by the EuroGEOSS project funded by the European Commission in the FP7 (Vaccari et al., 2012). Presently several FP7 projects adopts a brokering architecture; however, its maturity was demonstrated by GEOSS (Global Earth Observation System of Systems). In fact, the core component of the presently operative GEOSS Common Infrastructure (GCI) is the GEO DAB (Nativi et al., 2013): Discovery and Access Broker. This is a Brokering Framework providing discovery, evaluation and access services across the many heterogeneous capacities (i.e. archives, systems, disciplinary infrastructures, networks, etc.) which contribute to GEOSS in a very loosely-coupled way. More recently, the NSF Earth Cube programme has investigated the brokering technologies recognizing its importance for realizing global and multidisciplinary interoperability (The EarthCube Team, 2012).

The scope and long-term vision of NextData require global and multidisciplinary research, indeed. Therefore, it is necessary to design and implement a cyber(e)-infrastructure that applies the brokering approach principles (Nativi et al., 2012):

1. To keep the existing capacities as autonomous as possible by interconnecting and mediating between standard-based and non-standard-based capacities.
2. To supplement, without supplanting, the individual systems' mandates and governance arrangements.
3. To assure a low entry barrier for both the resource providers and the end users.
4. To be flexible enough so as to accommodate the existing systems as well as future ones.
5. To build in an incremental fashion upon the existing infrastructures (information systems) and incorporate heterogeneous resources by introducing distribution and mediation functionalities.

6. To specify interoperability arrangements focusing on the modularity of interdisciplinary concepts rather than just on the technical interoperability of systems.

In keeping with this principles and applying the brokering architecture successfully demonstrated by GEOSS, the NextData cyber(e)-Infrastructure will be structured as depicted in Figure 1.

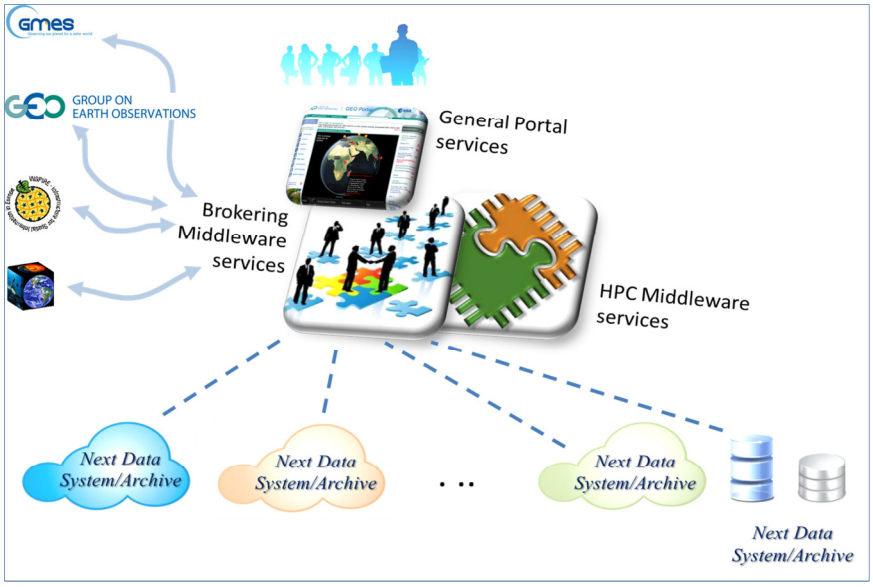

Figure 1. The NextData cyber(e)-Infrastructure

Building on the heterogeneous NextData systems and archives, the cyber(e)-infrastructure will broker all of them through the brokering framework middleware that provides the necessary intermediation and harmonization services and supplements those systems with the capabilities they are short of. The brokering framework, in turn, makes use of the high-performance computing services provided by a HPC middleware platform (e.g. cloud computing technology). Finally, the brokering framework is accessed by the NextData General Portal providing the graphical interface for the Human-Computer Interaction. All the brokering core and advanced features are exposed through the General Portal to the Users.

References

Nativi S., Craglia M., Pearlman J., 2013, *Earth Science Infrastructures Interoperability: the Brokering Approach,* in press on IEEE JSTARS.

Nativi S., Craglia M., Pearlman J., 2012, *The Brokering Approach for Multidisciplinary Interoperability: A Position Paper.* International Journal of Spatial Data Infrastructures Research, vol.7, pp. 1-15.

Nativi S., Fox P., 2010, *Advocating for the use of informatics in the earth and Space Sciences*, EOS, Transactions, American Geophysical Union, Vol. 91, No 8, pp. 75 – 76.

Vaccari L., Craglia M., Fugazza C., Nativi S., and Santoro M., 2012, Integrative Research: The EuroGEOSS Experience, IEEE JSTARS, vol. 5, n. 6, pp. 163-1611.

The EarthCube Brokering Concept Award Team, 2012, Brokering for EarthCube Communities: A Road Map, available at: http://api.ning.com/files/e1f7JW76fC4atlM7lc-wdrsh4bpyUb9knJhLXfEtBqjhhF*rgAdvZNTwzvtnWsqVSXA4EBbE6pNrBO*vpLKR1EiwePStjpnL/BrokeringRoadmapAug171.pdf

Milestones

Number	Description	Date
M1	**First prototype of the ND System of Systems infrastructure, implementing the core functionalities**	**12**
M2	**Consolidated version of the ND System of Systems infrastructure, implementing advanced functionalities**	**22**
M3	**Final and operative ND System of Systems infrastructure,**	**28**

Work Plan

ND-SoS-Ina is organized in five work-packages with core activities concentrated in WP2, WP3 and WP4. Both partners are, at different level, involved in the core WPs while only the coordinator CNR will be mainly involved in the management of the project and the dissemination activities.
Regular meetings (both virtual and in-person) and frequent contacts between the WP leaders will ensure the progress of the activities and the correct performance of the project in terms of timing and consumed resources.

- The management of the consortium, detailed in the WP1 description, is a continuous activity that starts at PM1.
- WP2 covers all the activities related to the collection of user requirements and the design of the system.
- WP3 represents the core of the project, focusing on the development, the integration of existing components and the testing of the system.
- WP4 is responsible for the operation of the development and production environment where the final system will be deployed.
- WP5 concerns the dissemination of project achievements and results.

Gantt chart

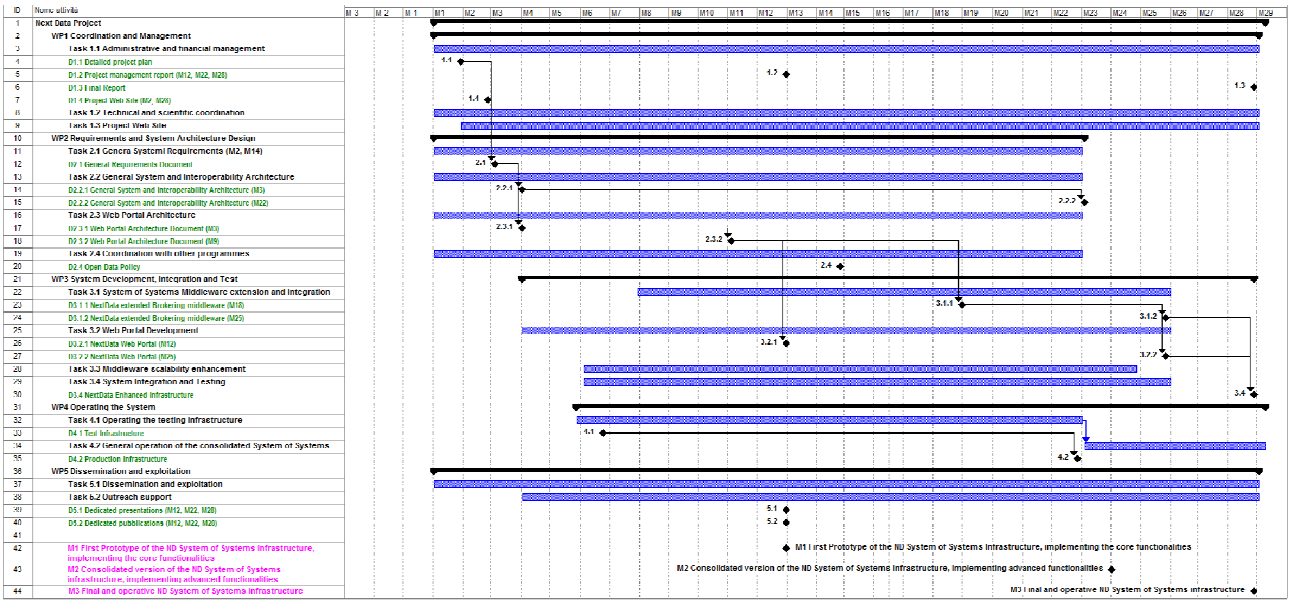

Figure 2. Gantt chart of the project activities.

WP	Title	Type	Lead	PM	Start month	End month
1	Coordination and management	Management	CNR	5	1	28
2	Requirements and System Architecture design	R&D	CNR	12	1	28
3	System Development, Integration, and Test	R&D	CINECA	52	2	28
4	Operating the System	Operation	CINECA	25	2	28
5	Dissemination and Outreach	Dissemination	CNR	11	1	28
			Total	105		

WP1 Coordination and Management (lead: CNR)				
Start Date: M1	**End Date: M28**	**Effort (person month)**		
		2013	**2014**	**2015**
		6	**6**	**7**
		Total Effort		
		CNR (2013+2014+2015)	**CINECA (2013+2014+2015)**	
		18 (6+6+6)	**1 (0+0+1)**	

Task 1.1 Administrative and financial management [M1-M28] [CNR]
This task will initiate, set and implement the consortium contract, including the

consortium agreement, the overall IP financial and administrative management processes and routines. The task lasts for the entire duration of the project and provides full control of all financial and administrative aspects of the IP.

Task 1.2 Technical and scientific coordination [M1-M28] [CNR]
The purpose of having a technical and scientific task inside the management Work Package is to remark the importance of Technical and Scientific coordination. These activities are:

- To ensure the coordination of technical and scientific exchanges between Work Packages.
- To assess all technical and scientific deliverables regarding their fitness for purpose of the overall technical and scientific goals.
- To assure coherence of all technical and scientific developments.
- To decide on changes or adaptations of technology employed in the project as a result of important recommendations from GEOSS, EarthCube, GMES and relevant International standards and specifications.
- To manage the technical risks.
- To coordinate the proposal activities with the NextData project leadership and general objectives.

Task 1.3 Project Web site [M2-M28] [CNR, CINECA]
This task will set up a light project web site and manage it throughout the project. The public part of the web site will be used for presenting the project results. An internal partner area will be used for exchanging project-internal information and for collaboration between the partners.

Deliverables:
D1.1 Detailed project plan (M1)
D1.2 Project management report (M12, M22, M28)
D1.3 Final Report (M28)
D1.4 Project Web site (M2-M28)

WP2 Requirements and System Architecture Design (lead: CNR)

Start Date:	End Date:	Effort (person month)		
		2013	2014	2015
M1	M28	17	12	8

Total Effort	
CNR (2013+2014+2015)	CINECA (2013+2014+2015)
32 (15+10+7)	5 (2+2+1)

Task 2.1 General Requirements [M1-M22] [CNR, CINECA]
It is a fundamental goal of this project to implement a system that user communities indeed need. To foster this major goal, this task will identify the user requirements in the functional

and non-technical areas based on their needs captured in the form of detailed use and service cases, identifying the adequate tools for data management, processing and visualization. Besides it will establish requirements for the operation of services, i.e. addressing the questions on expected level of availability, reliability, scalability and performance. Operational parameters will feed into the definition of the operational model and procedures for the system as well as the design of the hosting infrastructure (WP4).
User requirements will be further refined and customized according to the feedback from the use of the first prototype architecture and they will be exploited for its consolidated version.

Task 2.2 General System and Interoperability Architecture [M1-M22] [CNR, CINECA]

This task produces the overall architecture of the final system with respect to internal requirements and standards, or protocols, being produced by external activities or programmes.

Task 2.3 Web Portal Architecture [M1-M22] [CINECA, CNR]

This task is concerned with the definition of the Web Portal Architecture considering function and non-functional requirements being collected by task 2.1. A technology appraisal of existing technologies (EuroGEOSS, etc.) will precede the final design of the portal to identify which existing solutions and tools could be employed during the implementation. The documents produced by this task will drive the development of the portal and will be kept updated throughout the project.

Task 2.4 Coordination with other programmes [M1-M28] [CNR]

- European (INSPIRE, ESFRI, etc)
- International (GEO GEOSS, WMO, EarthCube, etc.)

This task will take care to establish a policy for the attribution of the Open Data in coordination with other programmes and with the recommendation of the European Commission (http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf).

In particular, the data portal will carefully consider: transparency, open government and innovation. It will provide access to open public data from NextData. It will also provide access to data of other projects, infrastructures, and agencies at their request. The published data can be downloaded by everyone interested to facilitate reuse, linking and the creation of innovative services. Moreover, the data portal will promote literacy about published data. The Open Data Access guidelines, under definition by Science Europe, will be considered and applied.

Deliverables:

D2.1 General System Requirements [M2 and M14];
D2.2 General System and Interoperability Architecture [M3 and M22];
D2.3 Web Portal Architecture [M3, M9]
D2.4 Open Data policy [M14]

<table>
<tr><th colspan="3">WP3 System Development, Integration, and Test (lead: CINECA)</th></tr>
<tr><td>Start Date:
M2</td><td>End Date:
M25</td><td>Effort (person month)
<table>
<tr><th>2013</th><th>2014</th><th>2015</th></tr>
<tr><td>31</td><td>28</td><td>21</td></tr>
</table>
Total Effort
<table>
<tr><th>CNR
(2013+2014+2015)</th><th>CINECA
(2013+2014+2015)</th></tr>
<tr><td>47 (17+18+12)</td><td>33 (14+10+9)</td></tr>
</table>
</td></tr>
<tr><td colspan="3">Task 3.1 System of Systems middleware extension and integration [M8-M25] [CNR]
This task includes the following objectives:
• Brokering middleware (GEO DAB technology) Integration with the portal and development of missing stubs, connectors.
• promoting the adoption of a system of controlled vocabularies (thesauri) currently developed by the EKOLab research unit of CNR-IIA for the interoperable management of environmental information within the different NextData archives.
• implementing a middleware to manage geospatial data and metadata; this will also support (where necessary) data acquisition from monitoring stations and Quality Assurance/Quality Control processes following the large experience of the global monitoring network on mercury.
Task 3.2 Web Portal development [M3-M25] [CINECA]
This task pertains with the development of the Web Portal. The implementation will be organized on two different phases to accelerate the delivery of developed functionalities and ensure an incremental development of user requirements.
• First phase [M3-M12]
Based on the outcomes of WP2, during this phase, the task will utilize available technologies and tools to implement a first version of the web portal providing the most relevant functionalities. The portal will build upon the services and interfaces being offered by the brokering middleware.
• Second phase [M14-M25]
The second phase will focus on expanding core functionalities of the web portal, considering either further outcomes from WP2 or feedbacks from users based on the first release.
Task 3.3 Middleware scalability enhancement [M6-M24] [CINECA/CNR]
• Improving Access Broker middleware scalability;
• Improving Discovery Broker middleware scalability.
Task 3.4 System Integration and Testing [M6-M25] [CNR/CINECA]
• Integrate the developed software components (e.g. Web Portal, Brokering middleware, semantic assets, etc.) and run integration tests.</td></tr>
<tr><td colspan="3"><u>Deliverables</u>:
D3.1 NextData extended Brokering middleware [M18 and M25];
D3.2 NextData Web Portal [M12 and M25];</td></tr>
</table>

D3.3 NextData enhanced infrastructure [M27]

WP4 Operating the System (lead: CINECA)				
Start Date: M2	**End Date: M28**	**Effort (person month)**		
		2013	**2014**	**2015**
		10	**14**	**20**
		Total Effort		
		CNR (2013+2014+2015)	**CINECA (2013+2014+2015)**	
		24 (5+7+12)	**20 (5+7+8)**	

Task 4.1 Operation of the testing infrastructure [M2-M22] [CINECA]

This task will deploy, operate and maintain the system on an infrastructure satisfying the requirements from WP2 and able for testing either the services or requirements such as reliability, scalability, etc. CINECA will make available the resources being necessary for the development, the deployment and the operation of the final system.

Task 4.2 General operation of the consolidated System of Systems [M22-M28]

- **General operation of the Web Portal [CNR]**
- **General operation of the infrastructure [CINECA]**
  The system of system will operate on a production infrastructure to guarantee the requirements such as availability, reliability, scalability, monitoring, security etc.

Deliverables:

D4.1 Test infrastructure [M5];

D4.2 Production infrastructure [M22];

WP5 Dissemination (lead: CNR)				
Start Date: M1	**End Date: M28**	**Effort (person month)**		
		2013	**2014**	**2015**
		10	**13**	**13**
		Total Effort		
		CNR (2013+2014+2015)	**CINECA (2013+2014+2015)**	
		34 (10+12+12)	**2 (0+1+1)**	

Task 5.1 Dissemination and exploitation [M1-M28] [CNR, CINECA]

This task includes the following activities:

- Raise awareness of solutions and best practices developed in the program among the European and International communities working in the project application areas.
- Support development of international collaboration in the project areas with the objectives of creating opportunities for expanding to a global information base.
- Support outreach to the broad technical community, with emphasis on developing countries, through web-based and in-person information on the tools and standards and other technical developments of the program.

Task 5.2 Outreach support [M3-M28] [CNR]
Working closely with the NextData partners, this task will identify and facilitate opportunities for global participation in the NextData application areas. This includes meetings at international locations, seminars and smaller meetings in collaboration, wherever possible, with the GEO Communities of Practice. Furthermore, it will advise the project management on outreach opportunities.

Deliverables:
D5.1 Dedicated presentations [12, 22, 28]
D5.2 Dedicated publications [12, 22, 28]