Project of Strategic Interest NEXTDATA

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

Deliverable D2.7.6a
Open Data Policy

WP Coordinator: Stefano Nativi
CNR – IIA

Document identifier:	D2.7.6.a
Date:	Apr 2015
Editors:	Stefano Nativi, Enrico Boldrini, Fabrizio Papeschi
Activity:	SoSIna WP2 - Task 2.1 - 2.2 - 2.4
Lead Partner:	CNR
Document status:	Final
Document link:	http://essi-lab.eu/nextdata/sosina/?page_id=68
Diffusion	Internal
Abstract	This document presents the principles for Data Management and Data Sharing that NextData resource providers shall implement to assure effective and sustainable interoperability and contribute to the NextData general infrastructure.
Keywords	Open Data, Data Access, Data Management, Interoperability .

History

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

Abstract

This document presents the principles for Data Management and Data Sharing that NextData resource providers shall implement to assure effective and sustainable interoperability and contribute to the NextData general infrastructure.

Keywords

Open Data, Data Access, Data Management, Interoperability.

TABLE OF CONTENTS

1. Introduction

1.1 Broker pattern: The NextData system of systems

The NextData system infrastructure implements a fully distributed system addressing remote system heterogeneity through the Broker pattern.

The **Broker pattern** separates users of services (NextData users) from providers of services (NextData data providers) by inserting an intermediary, called a broker.
When a user needs a service, it queries a broker via a service interface. The broker then forwards the user's service request to one or more data providers, which processes the request.

The implementation of a brokering framework allows to implements other important software engineering principles, as: indirection, delegation, low coupling, protected variations, etc.

The NextData system infrastructure will apply the following distribution principles, characterizing the System of Systems (SoS) approach –see for example the GEOSS Common Infrastructure (GCI) [R1]:
- A. Broker existing system/network infrastructures.
- B. Supplement but not supplant brokered system/network responsibilities (e.g. mandates and governance arrangements).
- C. Address brokered system heterogeneity harmonizing their information through a set of mediation services.
- D. Be flexible and extensible to:
    - i. Broker new and diverse system/network infrastructures.
    - ii. Sustain and advance the achieved interoperability.
    - iii. Allow each remote system/network (i.e. Community) infrastructure to evolve.

1.2 Expert pattern: NextData information content management

In a fully distributed SoS (like NextData), every request submitted by users is forwarded by the broker to the appropriate set of brokered data systems. Where required, the broker translates the request encoding to match each remote server data model, decoupling users and data providers. This applies the well-known Expert pattern.

Information Expert (also **Expert pattern**) is a principle used to determine where to delegate responsibilities. Using the principle of Information Expert, a general approach to assigning responsibilities is to look at a given responsibility, determine the information needed to fulfill it, and then determine where that information is stored. Information Expert will lead to placing the responsibility on the entity (e.g. service) with the most abundant information required to fulfill the task.

In extreme summary, the NextData information content (i.e. data and metadata) is generated and managed by the remote data providers. The NextData broker component is an instrument to decouple resources, harmonize contents and protocols, transform and encoding content and assess it, but it cannot supplement information that is missing from data providers.

2 Responsibilities for users' request finalization

Therefore, considering a given user's request, these are the main responsibilities:
- **Broker responsibilities**:
    1. General validation of the request.
    2. Mediation (translation of the request encoding to match each remote server data model).
    3. Supplementation of the Brokered systems, where necessary –e.g. data simple transformation.
- **Data Providers responsibilities**:
    1. Provision of the information necessary to finalize the request –e.g. discovery, access, etc.

3 NextData general data (sharing) policy

The general NextData data policy introduced the following sharing principles:
- Full and open exchange of data, metadata and products within the NextData Project, 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.
- Metadata will be made publicly available and data obtained with Project funding will be made available by the end of the Project.
- It will be encouraged to provide all shared data, metadata and products free of charge or at no more than the cost of reproduction for research and education.

4 Open Data Management and Sharing principles

4.1 Open Data Management principles for NextData information providers

To effectively respond to the NextData users' requests, each data provider will implement a set of Data Management principles aiming to guarantee a standard of quality for data discovery and accessibility. These principles are taken from the GCI Data Management principles, recently proposed by the GEOSS IIB (Infrastructure Implementation Board) [R2].

4.1.1 Discoverability
1. Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions will be clearly indicated.

4.1.2 Accessibility
2. Data will be openly accessible with minimum delay and cost.

3. Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.

4.1.3 Usability

4. Data will be structured on open-standards and encodings that are harmonized to the greatest extent possible given organizational needs and observing methods.
5. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international standards.
6. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products.
7. Data will be quality-controlled and the results of quality control will be indicated in the metadata; data made available before quality control will be flagged in metadata as unchecked.

4.1.4 Preservation

8. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.
9. Data and associated metadata will be periodically verified to ensure integrity, authenticity and readability.

4.1.5 Curation

10. Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate.
11. Data will be assigned persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement for the use of their data.

4.2 NextData information Providers DMP

In addition to the general principles characterizing the NextData data sharing policy (see paragraph 3), in keeping with the H2020 European Commission requirement, each Data Provider will implement a Data Management Plan, applying a set of good practices [R3].

4.2.1 Data Management Plan

Each Data Provider will share a Data Management Plan (DMP) that will be kept updated during all the Project duration.
For each (external and or internal) resource type and the related software, the DMP will consider the following aspects:

Discoverability
- readily located;
- standard identification mechanism.

Accessibility
- modalities, scope, licenses.

Assessability and intelligibility
- scientific scrutiny.
- peer review.

Usability beyond the original purpose for which it was collected
- safely stored in certified repositories;
- long term preservation and curation;
- minimum software, metadata and documentation to make it useful.

Interoperability to specific quality standards
- standards for data annotation and data exchange;
- recombinations with different datasets from different origins.

In terms of Resource Documentation, the following elements will be considered by the DMP:

Resource reference and name
- Resource Identifier.

Resource description (Description of the resource that will be generated or accessed)
- its origin (in case it is collected);
- nature and scale;
- to whom it could be useful;
- whether it underpins a scientific publication;
- information on the existence (or not) of similar resources;
- possibilities for integration and reuse.

Standards and metadata
- Reference to existing suitable standards of the discipline.
- If these do not exist, an outline on how and what metadata will be created.

Resource sharing (Description of how resources will be shared)
- access procedures;
- embargo periods (if any);
- outlines of technical mechanisms for dissemination and necessary software;
- other tools for enabling re-use;
- definition of whether access will be widely open or restricted to specific groups;
- identification of the repository where resources will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).
- In case the resource cannot be shared: the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Archiving and preservation (including storage and backup)
- Description of the procedures that will be put in place for long-term preservation.
- Indication of how long the data should be preserved,

- its approximated end volume
- the associated costs and how these are planned to be covered.

5   *References*

R1	Brokering Approach: S. NATIVI et AL., IEEE JSTARS manuscript 2013.
R2	GEO Data Management Principles Task Force (DMSP-TF), GEOSS Data Management Principles, draft document, 2014.
R3	European Commission –DG Research and Innovation, Guidelines on Data Management in Horizon 2020 ver 1.0, Dec 2013.