



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

KICK-OFF MEETING

**First Focus: The Open cloud
aim and structure**

Mario Ciampi (CNR ICAR)

Rome, November 25, 2022



FOSSR

Fostering Open Science in Social Science Research
Innovative tools and services to investigate economic and societal change



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

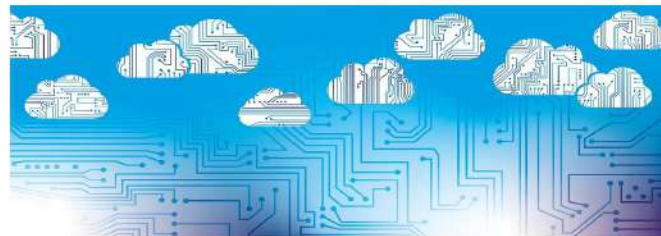


Consiglio Nazionale
della Ricerca

Objective

Creation of an *Italian SSH Cloud Computing and Data Infrastructure conforming to EOSC*

- *Single integrated system* of the Italian nodes of 3 European RIs and other data sources (like ISTAT)
- *Advanced tools and services* for a high quality analysis of research data





Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca



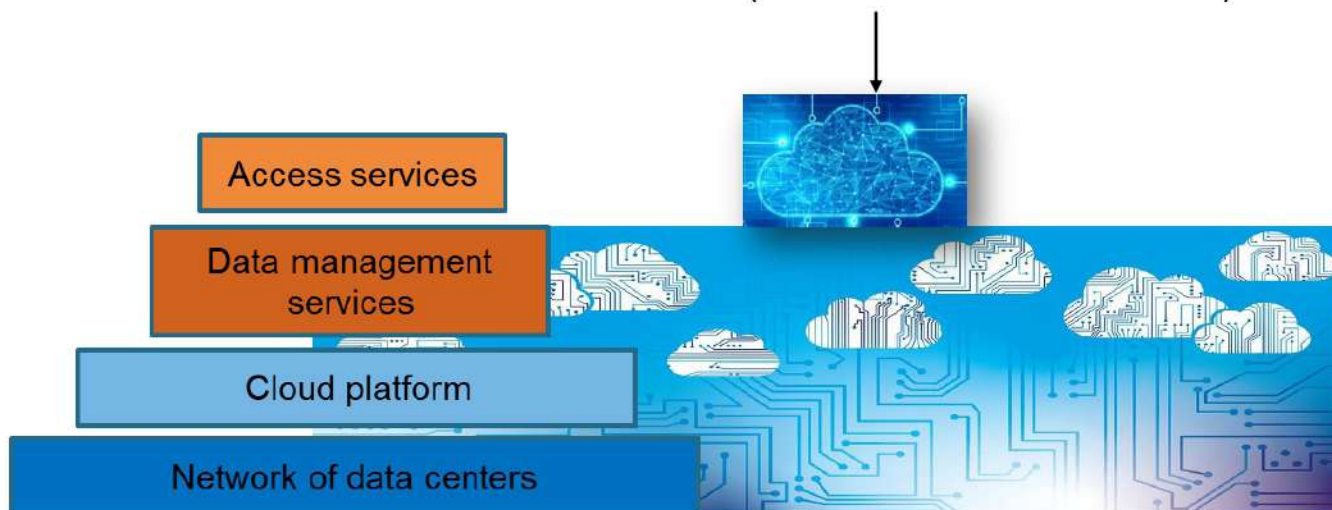
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Consiglio Nazionale delle Ricerche

Cloud Infrastructure Architecture

SINGLE POINT OF ACCESS (FEDERATED AUTHENTICATION)



 **cessda**
CONSORTIUM OF EUROPEAN
SOCIAL SCIENCE DATA ARCHIVES

RISIS 
RESEARCH INFRASTRUCTURE FOR SCIENCE
AND INNOVATION POLICY STUDIES

 **SHARE**
SURVEY OF HEALTH, AGEING
AND RETIREMENT IN EUROPE



Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

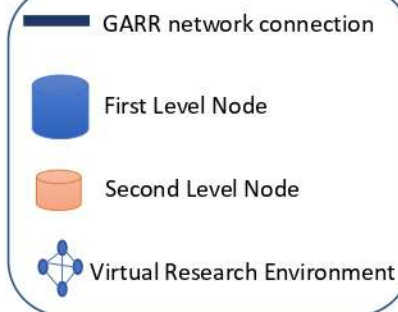


Consiglio Nazionale delle Ricerche

Network of data centers



Hardware and software resources for *high performance computing and big data storage* assuring *SLA* agreed with the existing RIs





Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca

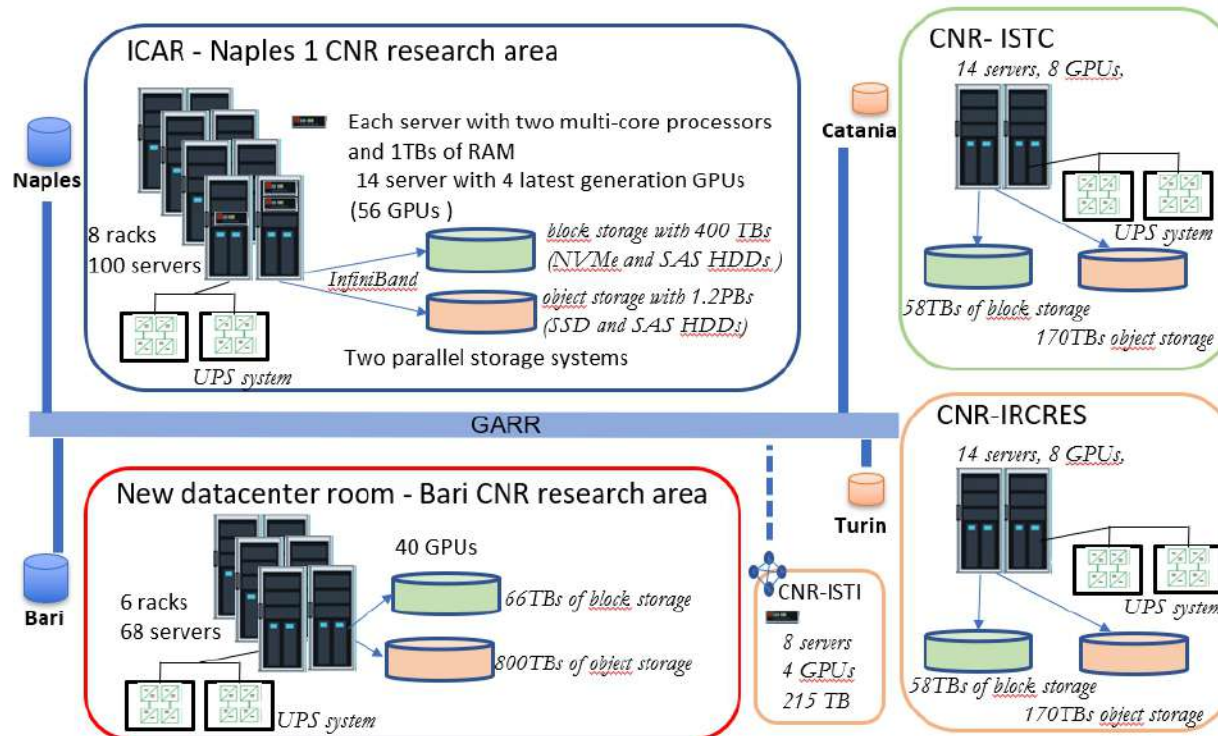


Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Consiglio Nazionale delle Ricerche

Data centers





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca

Cloud Platform

- *Federation* of the data centers in a single virtual system
- *Orchestrate* multiple composite applications
- Allocation of *virtual machines and containers*



Cloud Platform

according to

Efficient scheduling policies coherent
with the *DNSH principles*



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca

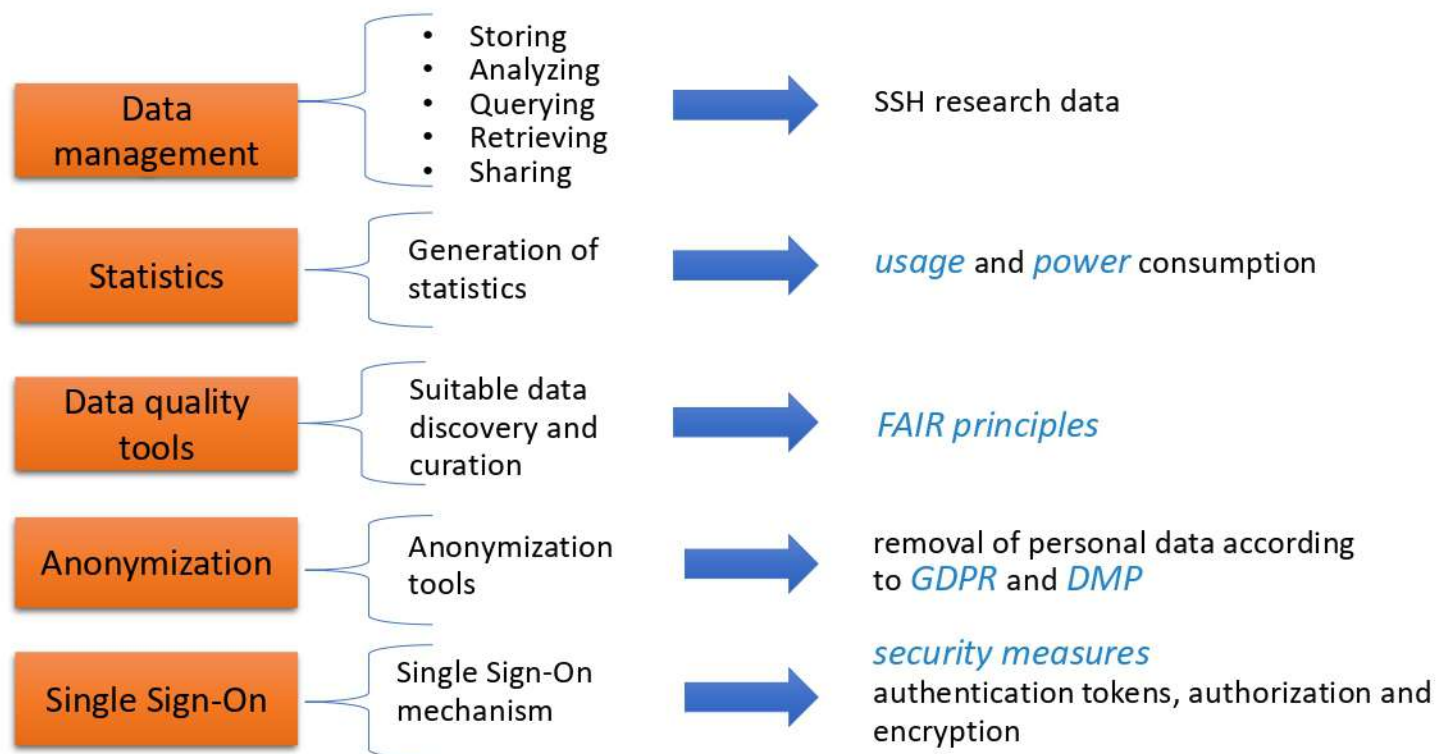


Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca

Data management Services





Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

Data management services Integration with national RIs

A unified **data model** will be defined

in order to

represent all the research open data handled by the existing RIs



CESSDA ERIC Consortium of European Social Science Data Archives European Research Infrastructure Consortium, of which Italy joined in July 2021, which provides the scientific community with facilities, tools, datasets, and certified services to conduct research activities of excellence in the social sciences domain.



SHARE ERIC Survey of Health, Aging and Retirement in Europe European Research Infrastructure Consortium, an interdisciplinary and longitudinal survey on the economic, social, health and well-being conditions of the 50+ population in twenty-seven European countries (plus Israel).



RISIS Research Infrastructure for Research and Innovation (R&I) Policy Studies, which provides data and services to support the development of a new generation of analyzes, and indicators for the study of science, technology and innovation processes on the base of three main perspective: actors involved to understand the role they play, topics addressed to understand the directionality of the R&I efforts, and geography of science and innovation.

The integration modalities will depend on each specific RI



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca

Data Quality

Data curation procedures

applied to available research data



Istat survey data



Data collected from
other involved
RIs/Universities



Data collected
during the project



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca

Respect of the FAIR principles

- Rich metadata description
- (Meta)data indexed in searchable resources

Findable

- Metadata always accessible
- (Meta)data retrievable by standardised protocols

Accessible

- Use of standard metadata
- Use of Controlled Vocabularies

Interoperable

- Data clearly licensed
- Data published in domain-relevant formats

Reusable



Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Consiglio Nazionale delle Ricerche

Access services

The open data managed by the open cloud will be accessible:

Marketplace (for humans)
Based on a multilingual web portal

Cloud Infrastructure



Host the resources and tools
Allow *access* for human users

Programmatic interfaces
(for applications)

REST-based APIs



defined and developed to ensure *interoperability*

Expressive query language

W3C standard SPARQL



solve the *data-silo problem*
ISTC
CNR



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche



Istituto di Scienza e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



D4SCIENCE

Virtual Research Environment (VRE)

A web working environment where Open Science practices are transparently promoted

Operated on top of ISTI D4Science Infrastructure featuring:

- Cloud storage area with **sharing functions**
- **Open Data analytics** platform
 - computational environment where **executions can be repeated, compared, discussed and logged**
 - Jupyter Notebooks via JupyterHub
- Catalogue enabling to **Find, Access** and **Reuse** every research artifact published from the VRE
- **Communication** platform via social applications



Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



Consiglio Nazionale delle Ricerche



Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



D4SCIENCE

Virtual Research Environment (VRE) at-a-glance

The image displays a Virtual Research Environment (VRE) interface, showing a dashboard with various components and a detailed view of a Jupyter Notebook.

Dashboard Components:

- Statistics:** A section showing user statistics and activity.
- News feed:** A section displaying recent news items, including a post about the 'VRE Factor for Coasts'.
- Blue-Cloud Project:** A section providing information about the project's goals and objectives.
- Blue-Cloud Lab:** A section detailing the lab's mission and the services it offers.
- Non-Protein Spaces:** A section describing the lab's focus on non-protein spaces and its research activities.
- Demonstrators:** A section listing various demonstrators, such as 'Bio-Phylogenetic ECV (Dem. 1)', 'Metagenomic Genomes (Dem. 2)', and 'Marine Env. Indicators (Dem. 2)'.

Jupyter Notebook View:

The notebook titled "Demonstrator 2 - Notebook 1.2. - Creating protein functional clusters for Notebook 2" includes:

- Introduction:** A text block discussing recent metagenomic studies and the challenges of analyzing large datasets.
- Notebook 1.2. allows for the creation of protein functional clusters from Fasta files derived from Metagenomic or Metatranscriptomic sequencing. These clusters contain annotated as well as unknown sequences and can then be passed on to Notebook 2.**
- Files for input:** A list of files available for use, including 'Tara Oceans Eukaryotic Genomes (the "EMAG")' and 'Metagenomic Genomes (EMAG)'.

Application Grid:

- JupyterLab:** A section describing the JupyterLab environment and its capabilities.
- RStudio:** A section describing the RStudio environment and its capabilities.
- Analysis Engine:** A section describing the Analysis Engine and its capabilities.
- Catalogue:** A section describing the Catalogue and its capabilities.



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca

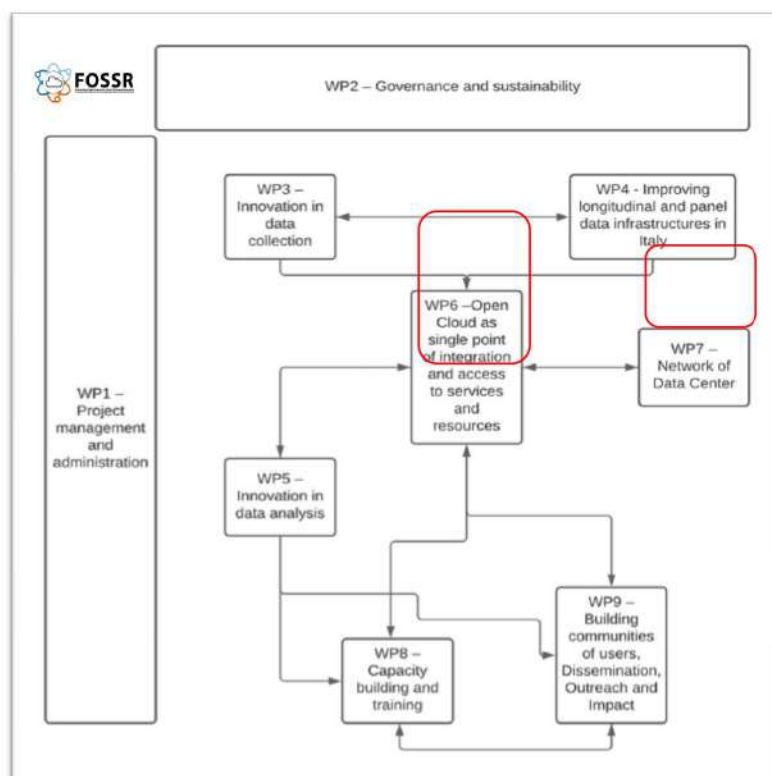


Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca

FOSSR Work Packages



Work Package 6: Open Cloud as single point of integration and access to services and resources

Leader: Mario Ciampi (CNR ICAR)
Senior Technologist

Budget: 3.870.933,65 €

Work Package 7: Network of Data Centers

Leader: Mario Sicuranza (CNR ICAR)
Technologist

Budget: 15.192.871,15 €



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

FOSSR Work Package 6

Task Nr.	Task Title	CNR Institute
6.1	Definition of user requirements, conceptual data model and system architecture of the platform	ICAR
6.2	Definition of the data management plan	ICAR
6.3	Definition of protocols for accessing and querying linked data in compliance with the FAIR principle	ISTC
6.4	Data discovery, curation, and anonymization	ICAR
6.5	Definition of the interoperability and security approaches for the access to resources and services	ICAR
6.6	Implementation of the cloud platform	ICAR
6.7	Implementation of the FOSSR web portal and tools	ICAR
6.8	Virtual Research Environment	ISTI



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
delle Ricerche

FOSSR Work Package 7

Task Nr.	Task Title	CNR Institute
7.1	Detailed design of the Data Centers	ICAR
7.2	Preparation, development and monitoring of tender procedures	ICAR
7.3	Naples first level node deployment	ICAR
7.4	Bari first level node deployment	ATR Bari
7.5	Catania second level node deployment	ISTC CT
7.6	Turin second level node deployment	IRCRES UIT
7.7	Installation of a cluster of servers in Pisa	ISTI
7.8	Integration tests, documentation and benchmarks	ICAR



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Consiglio Nazionale
della Ricerca



Any questions?