SESSIONE 7 / COSTI UMANI, COMPETENZE E SOSTENIBILITÀ ORGANIZZATIVA

Data Stewardship at CNR: Mapping a Professional Profile

Giorgia Miniello - CNR ITB | Claudio Lo Giudice - CNR IBIOM
[giorgia.miniello@cnr.it, claudio.logiudice@cnr.it]
per la Comunità dei Data Steward@CNR

GenOA week 2026 - Università di Genova, 2-5 Novembre 2026

Indice

- **The Data Steward (DS): a role at the intersection of research**
  - Definition of the Data Steward
  - Mapping activities and defining competencies

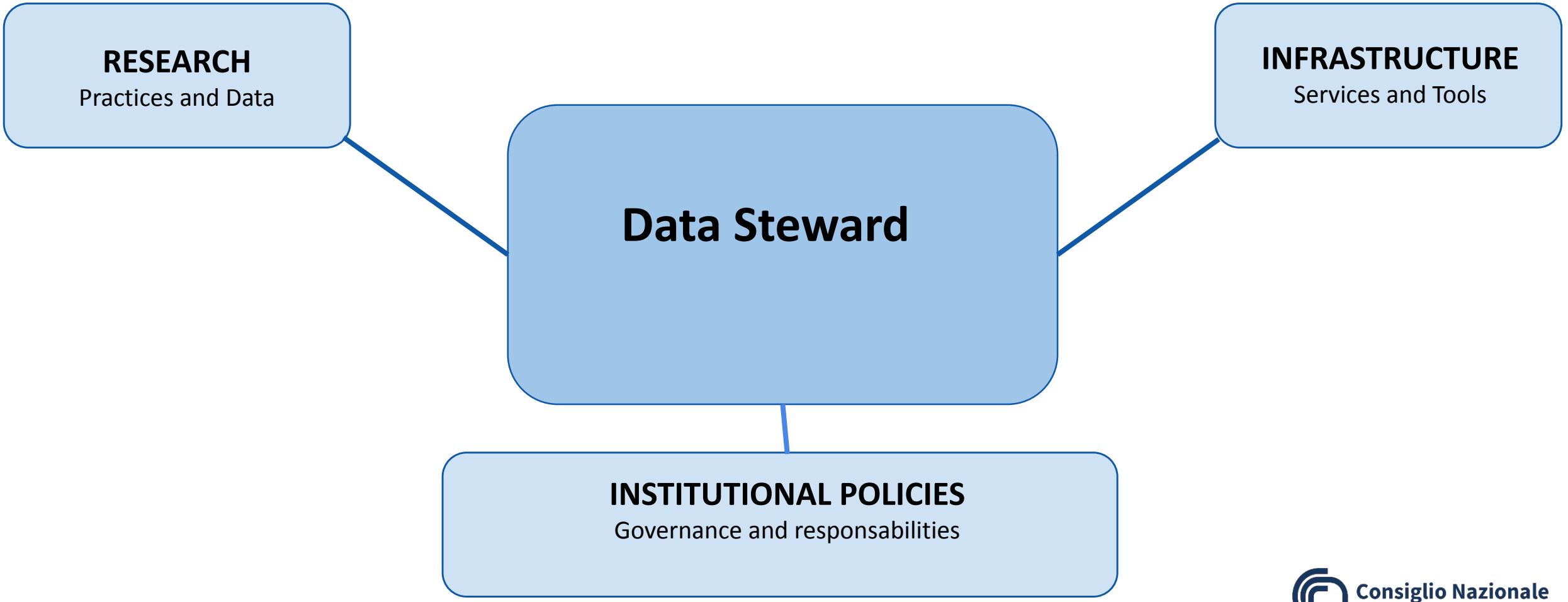
- **DS@CNR: the emergence of a community to be mapped**
  - Mapping methodology
  - *Self-selection sampling for Recruiting*
  - The Data Stewardship Activities Survey – DS@CNR as a structured self-assessment

- **Database DS@CNR: LIVE DEMO**
  - Main Dashboard and institutional profile
  - Statistics: CODATA keywords for monitoring activities
  - Geographical distribution

- **Conclusions**

The Data Steward: A Role at the Intersection of Research

FAIR principles, Data Management Plans and Open Science require multidisciplinary competencies.



Definition of a Data Steward

- According to the Italian Data Stewards Community:

> *"A Data Steward is a research data professional who supports researchers in the careful and responsible management of research data (and, more generally, all research outputs) throughout their lifecycle, from planning to public sharing, with the aim of fostering increasingly open and transparent science."*

Open Science

Source:
https://open-science.it/en/article?rpk=313614&prs_sel=p_datasteward&tpc_sel=t_gestione_dei_dati_della_ricerca

Mapping Activities and Defining Competencies at CNR

- Need to map activities and competencies related to Research Data Management (RDM).
- The Data Steward (DS) professional profile has not yet been formally institutionalized within CNR:
  - Contribute to **defining the DS professional profile based on the actual practices** of people working within the organization.
- **Objectives**:
  - Identify the key competencies associated with these roles and characterize their functions →**INSTITUTIONALIZATION.**
  - Identify potential **competency gaps and related training needs**→ **KNOWLEDGE SHARING AND TRAINING**

CNR Data Stewards' Vision Document (DOI: 10.57665/BICE_DATA2025)

DS@CNR: The Emergence of a Community to Be Mapped

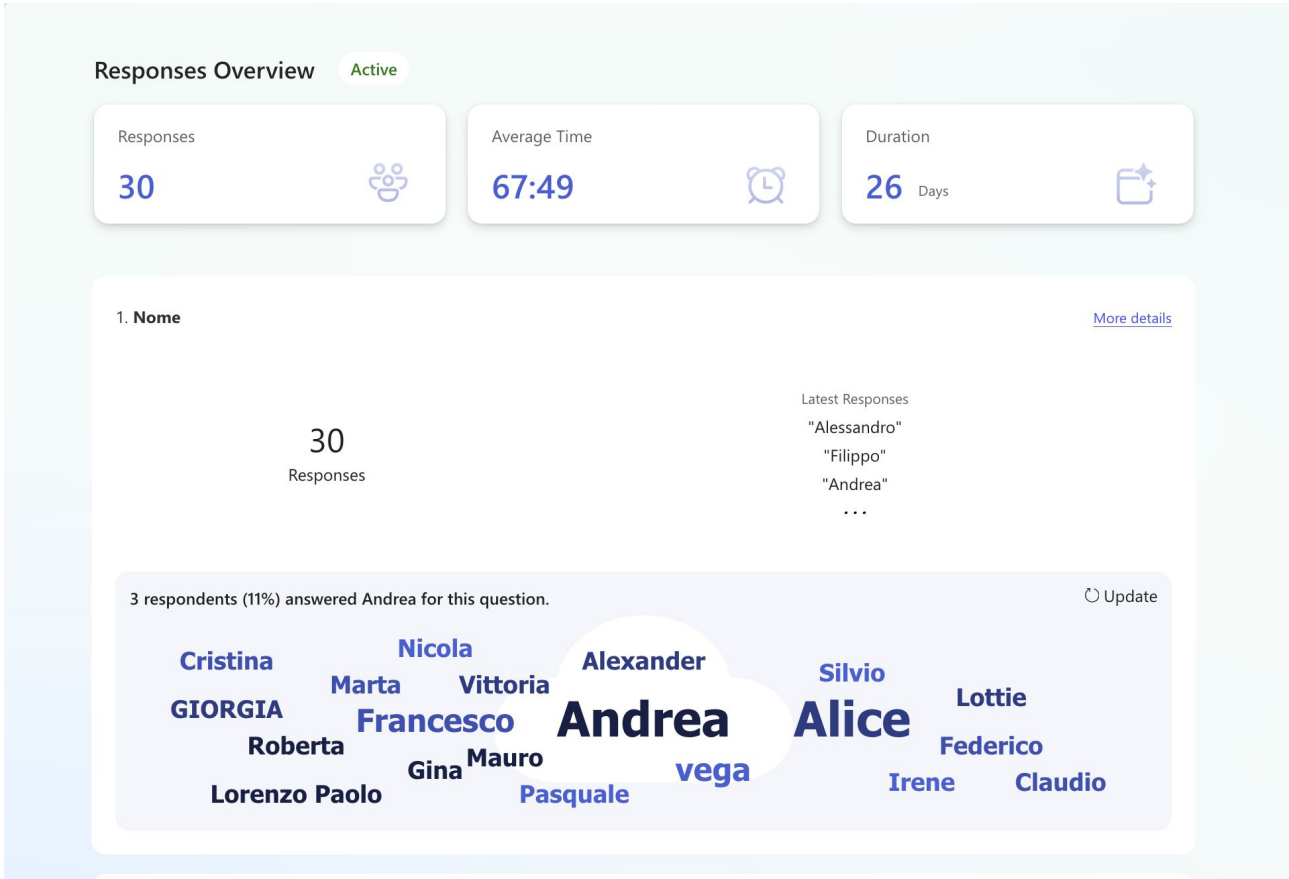
- Two-stage mapping methodology:
  - *Self-selection sampling for recruitment:* voluntary recruitment between December 2025 and May 2026, through institutional invitations sent to Department Directors for dissemination to interested individuals
    - Participants joined the **DS@CNR Community** (76 people)
    - shared Excel spreadsheets were used for initial activity mapping.

  - Data Stewardship Activities Survey – DS@CNR, organized into five main sections:
    - Personal and institutional profile
    - Role and employment classification
    - Education and background
    - Activities performed and future opportunities
    - Existing competencies, needs and training

Additional sections cover comments and suggestions, privacy, data processing and consent.

Data Stewardship Activities Survey – DS@CNR

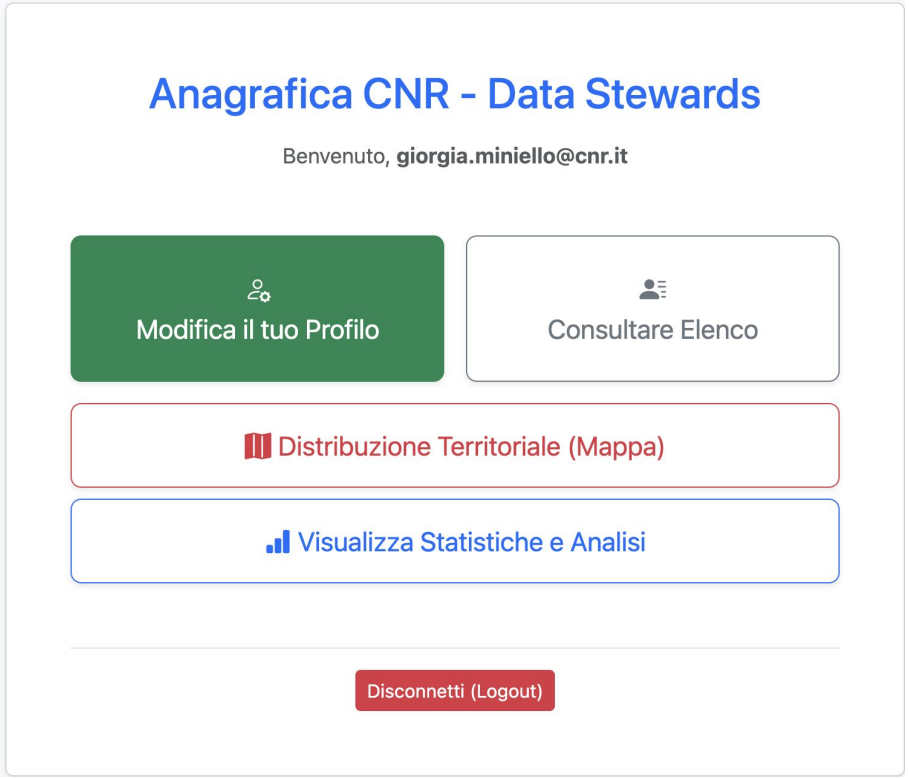


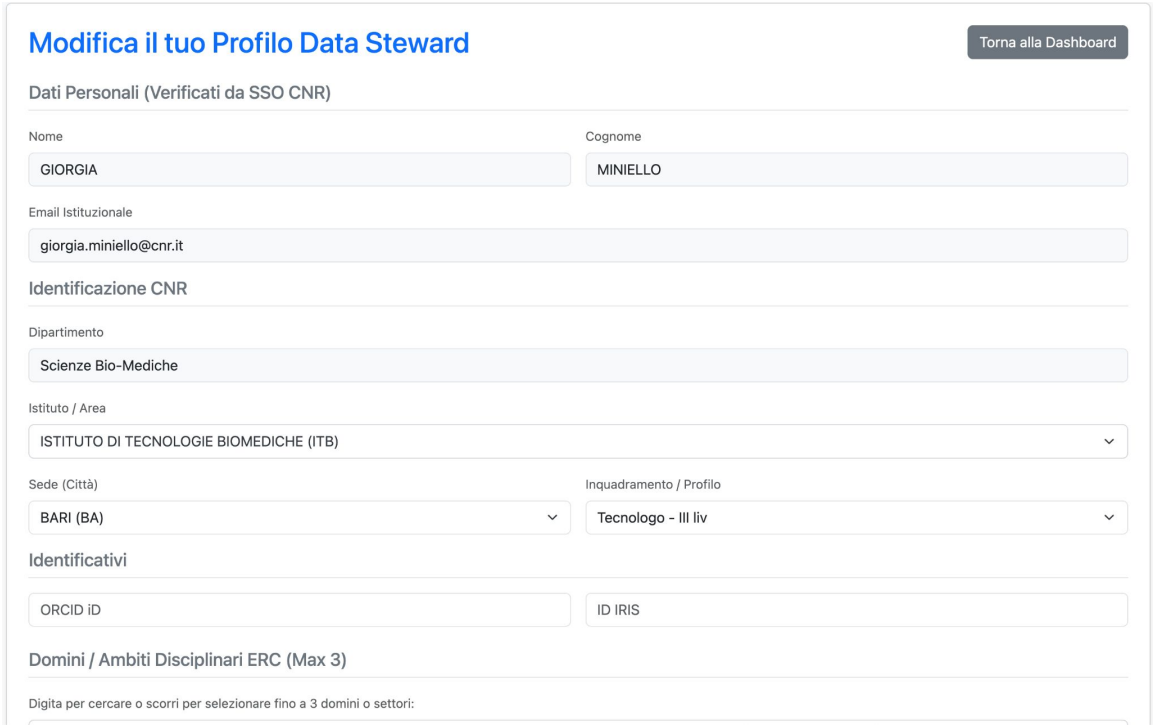


[Survey Attività di Data Stewardship - DS@CNR – Fill out form]

DS@CNR Database: Main Dashboard and Institutional Profile

- The survey data were collected in a dedicated **NoSQL document database**, accessible at https://dsdatabase.cnrbiomics.it/
- Access to the platform is enabled through institutional email credentials; integration with CNR Single Sign-On (SSO) is in progress.
- Main dashboard functions: Edit your profile; Browse the directory; Geographical distribution (map); View statistics and analyses.





DS@CNR Database: Technical Details

INFRASTRUCTURE

Technological Stack & Middleware

Flask Backend

Python microframework configured with ProxyFix behind an HTTPS reverse proxy.

MongoDB Store

Document database on a dedicated host, mongo-cnr (port 27017).

Authlib & OAuth

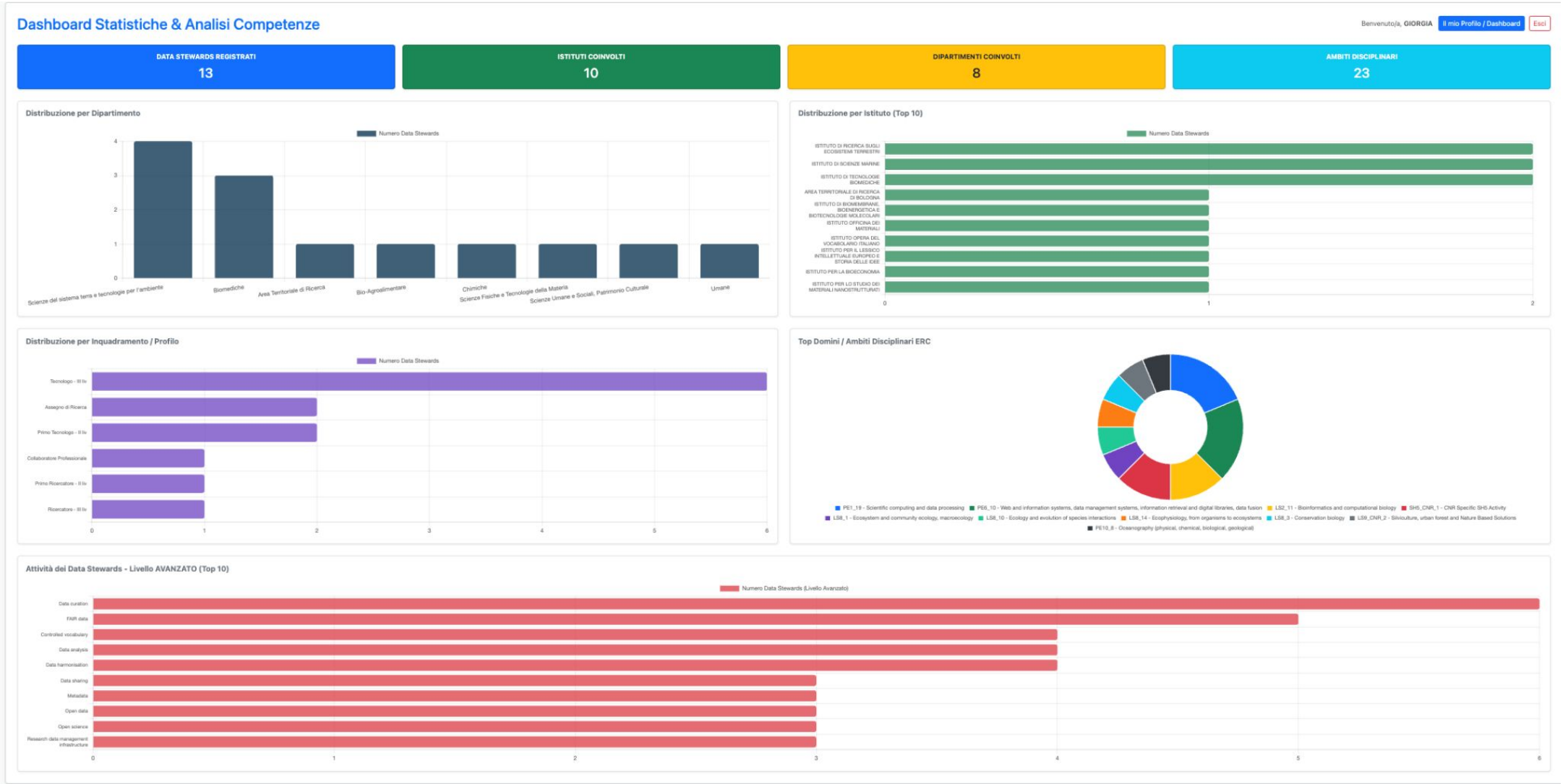
OIDC client integrated with the CNR Keycloak realm for secure SSO.

```
CONTAINER ID   IMAGE                              COMMAND                  CREATED       STATUS       PORTS                                                NAMES
53fdef345814   quay.io/keycloak/keycloak:24.0.1   "/opt/keycloak/bin/k…"   10 days ago   Up 10 days   8080/tcp, 8443/tcp                                   cnr-keycloak
99ccf6cbec20   nginx:alpine                       "/docker-entrypoint.…"   11 days ago   Up 11 days   80/tcp, 0.0.0.0:443->443/tcp, [::]:443->443/tcp      cnr-nginx
f501fc88499f   dswdb-app                          "python app.py"          11 days ago   Up 3 days    5000/tcp                                             dswdb-app
a9dff3b4f89c   postgres:15-alpine                 "docker-entrypoint.s…"   11 days ago   Up 11 days   5432/tcp                                             keycloak-db
e6a5e37968c8   mongo:8.2                          "docker-entrypoint.s…"   11 days ago   Up 11 days   0.0.0.0:27017->27017/tcp, [::]:27017->27017/tcp      mongo-cnr
```

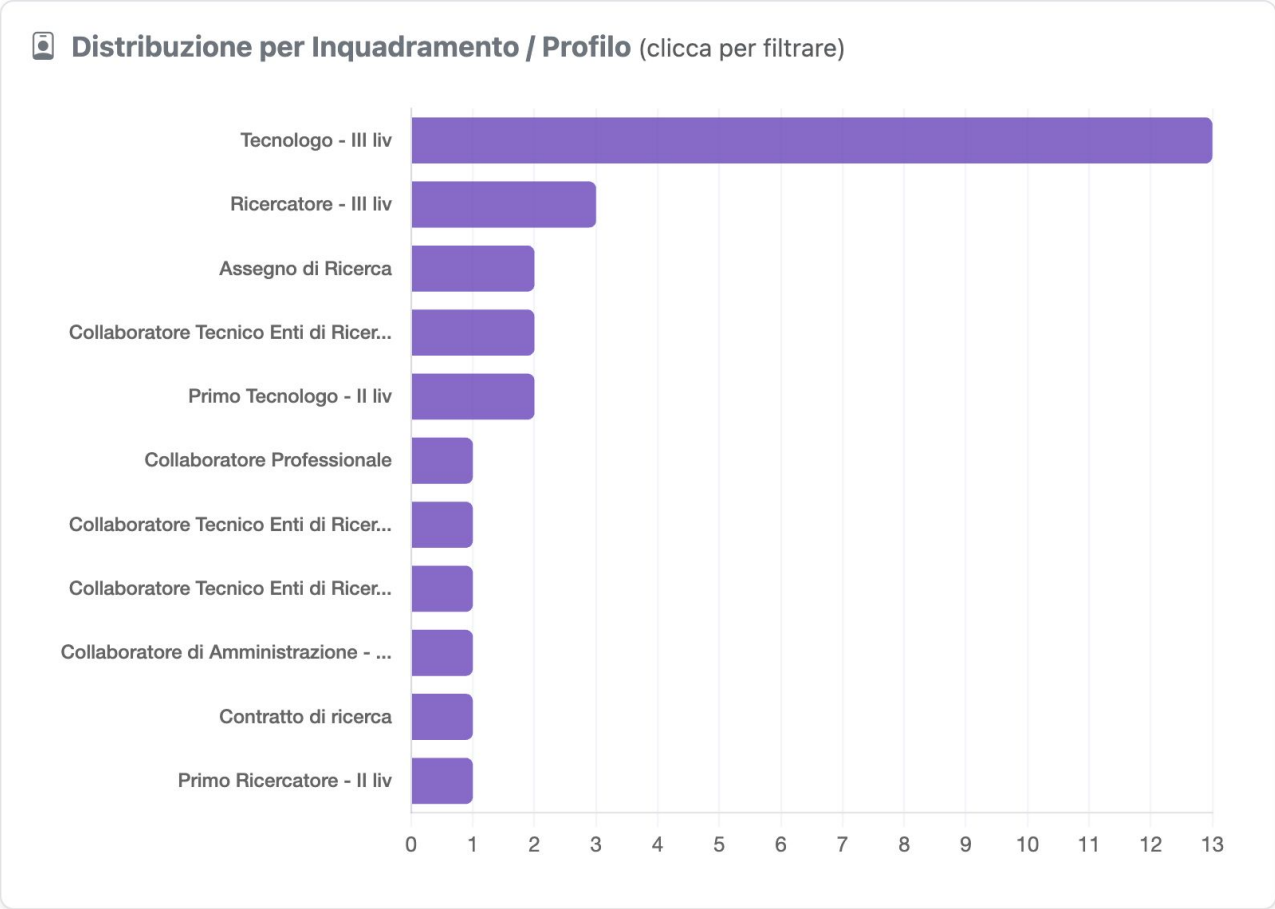
LIVE DEMO - DSDATABASE@CNR

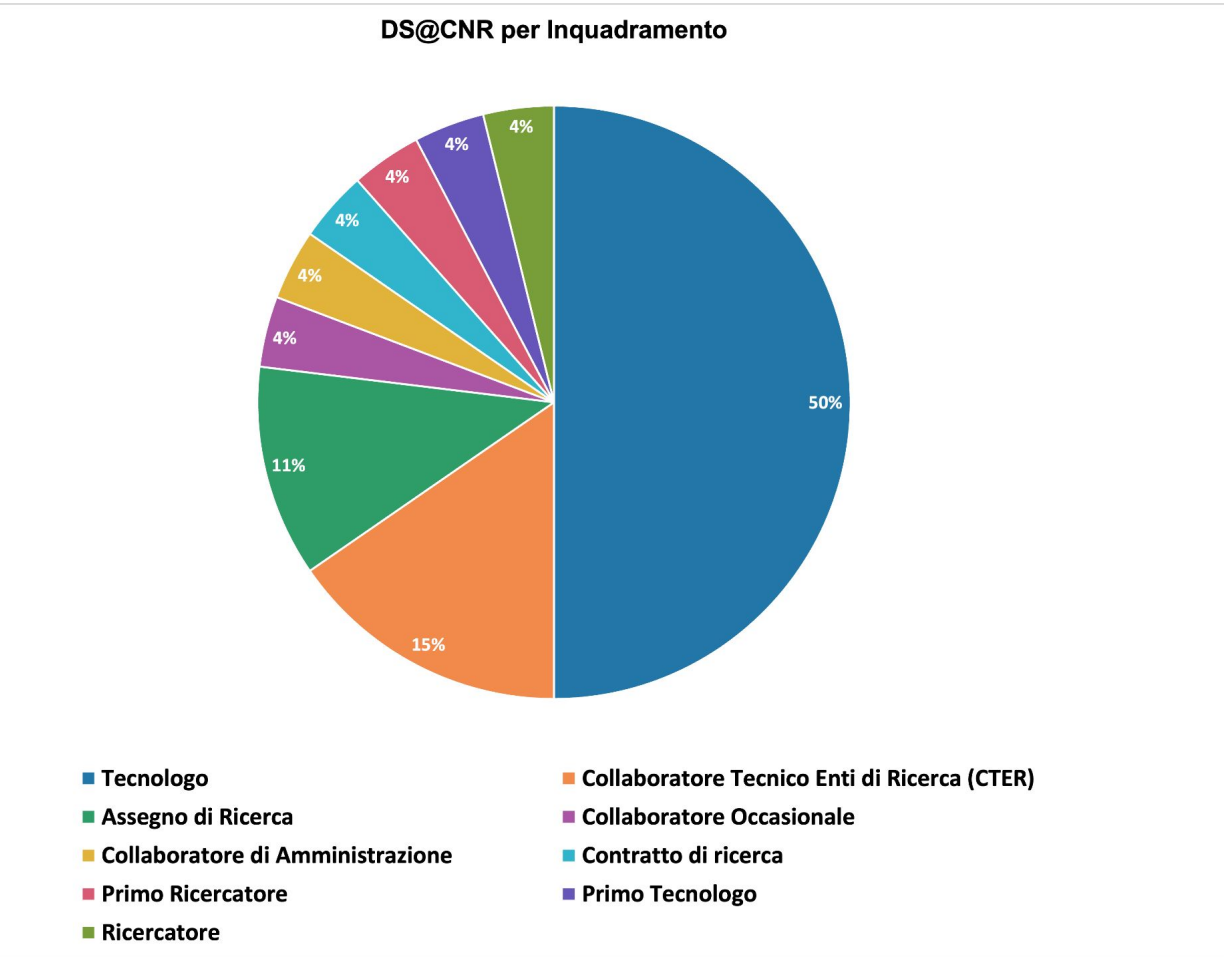
Live demonstration of the **DS@CNR database** and its interactive features.

DS@CNR Database: Statistics Dashboard [⅕]

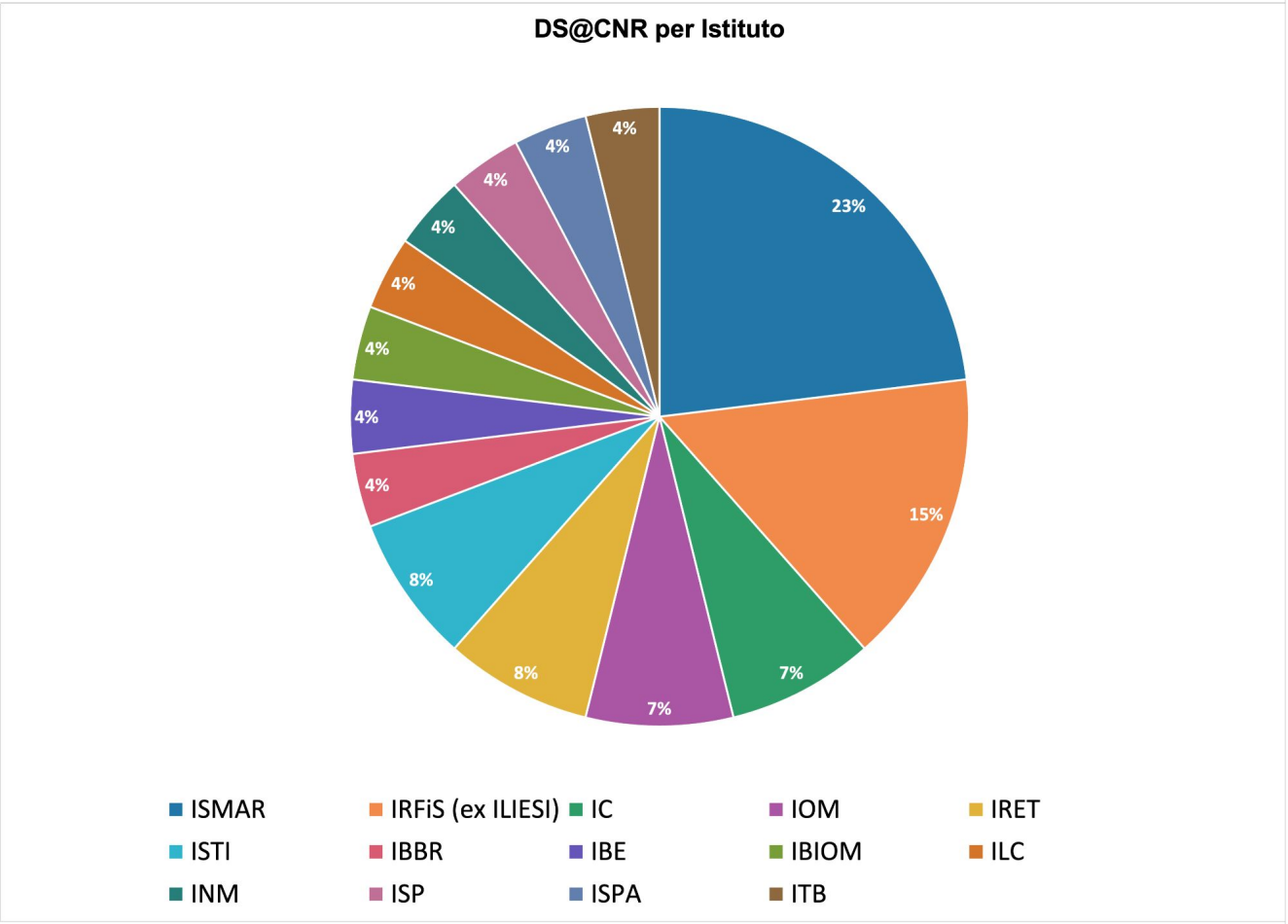


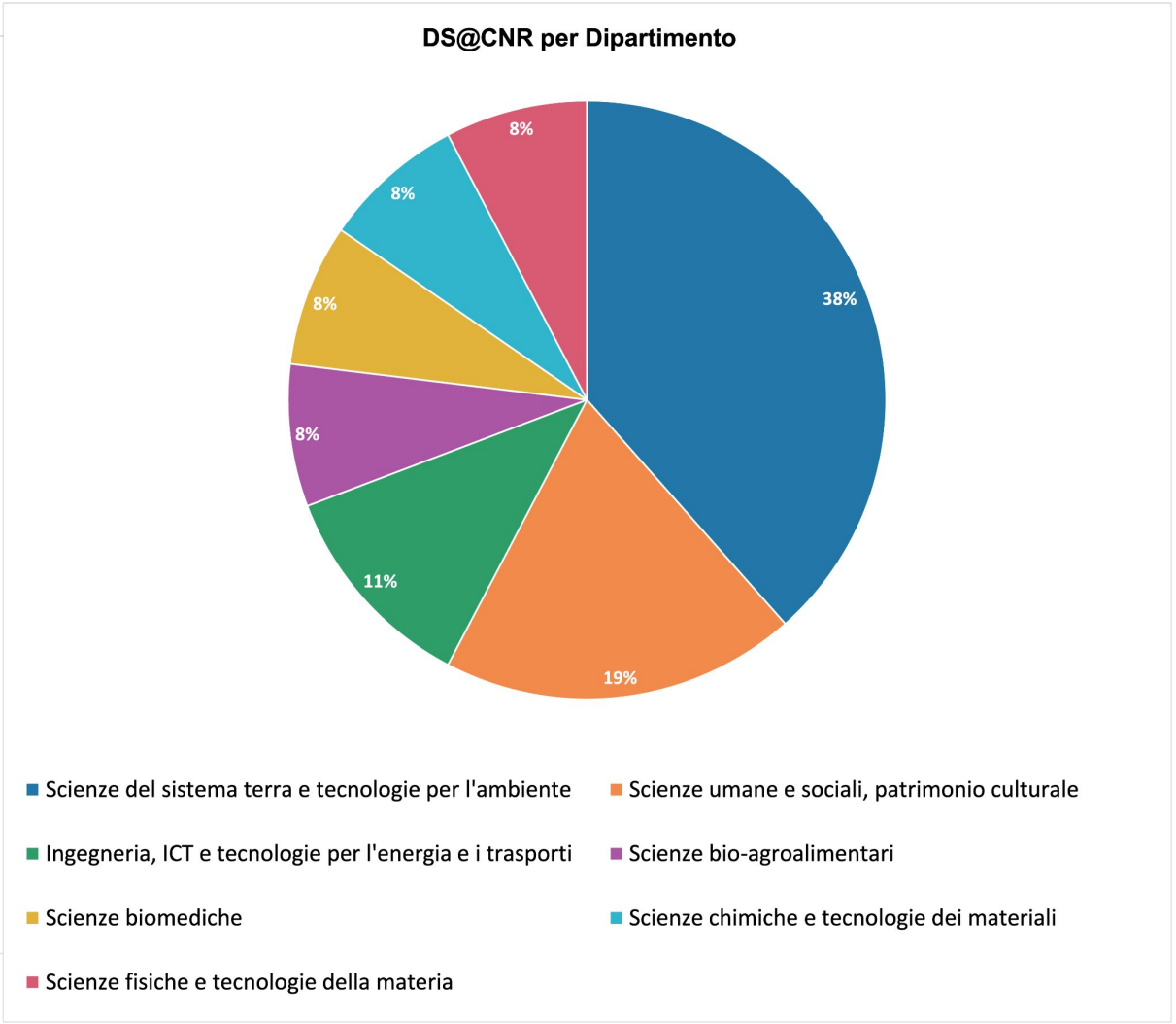
DS@CNR Database: Professional Profile Statistics [⅔]



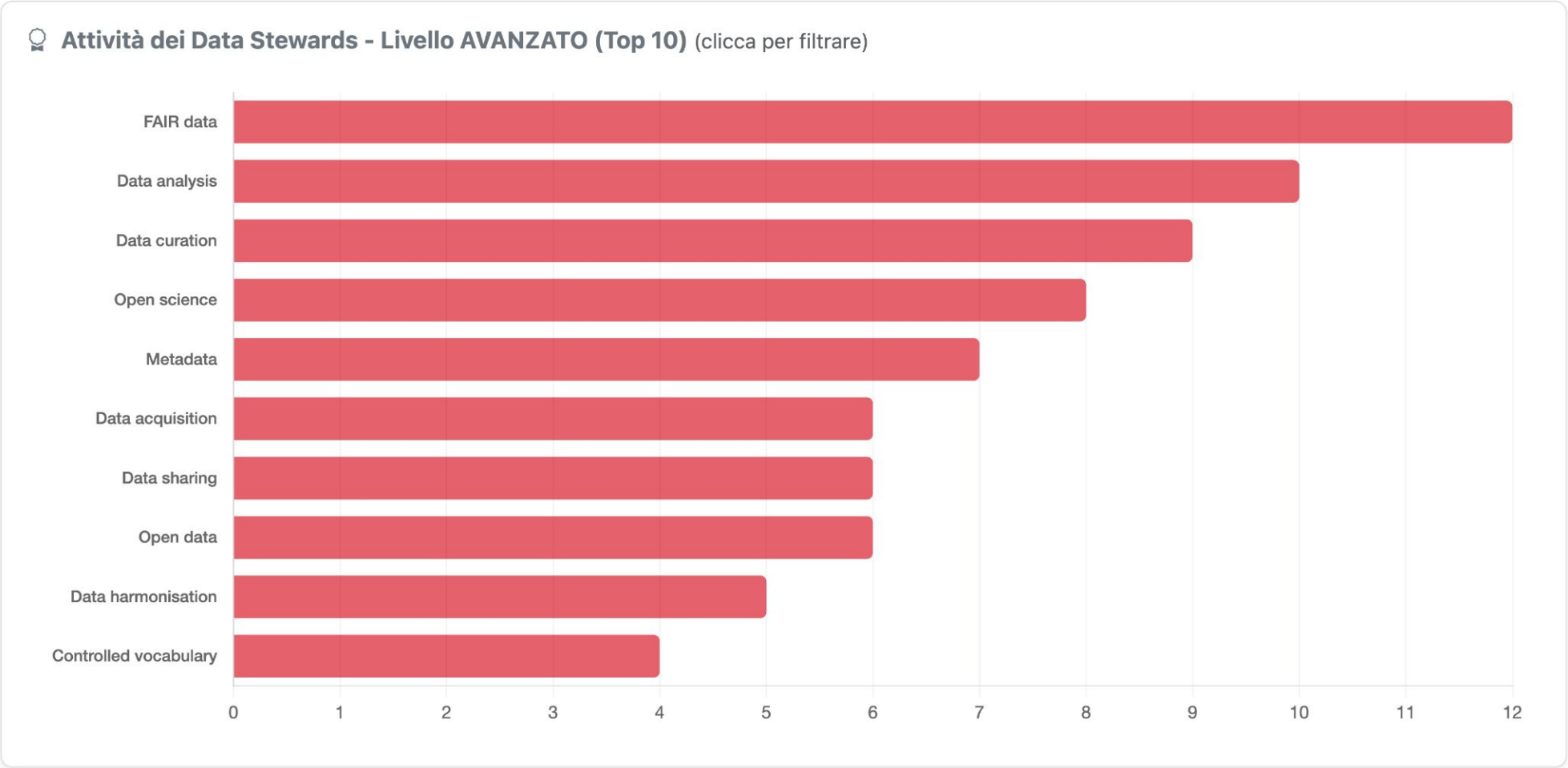


DS@CNR Database: Institute and Department Statistics [3/5]



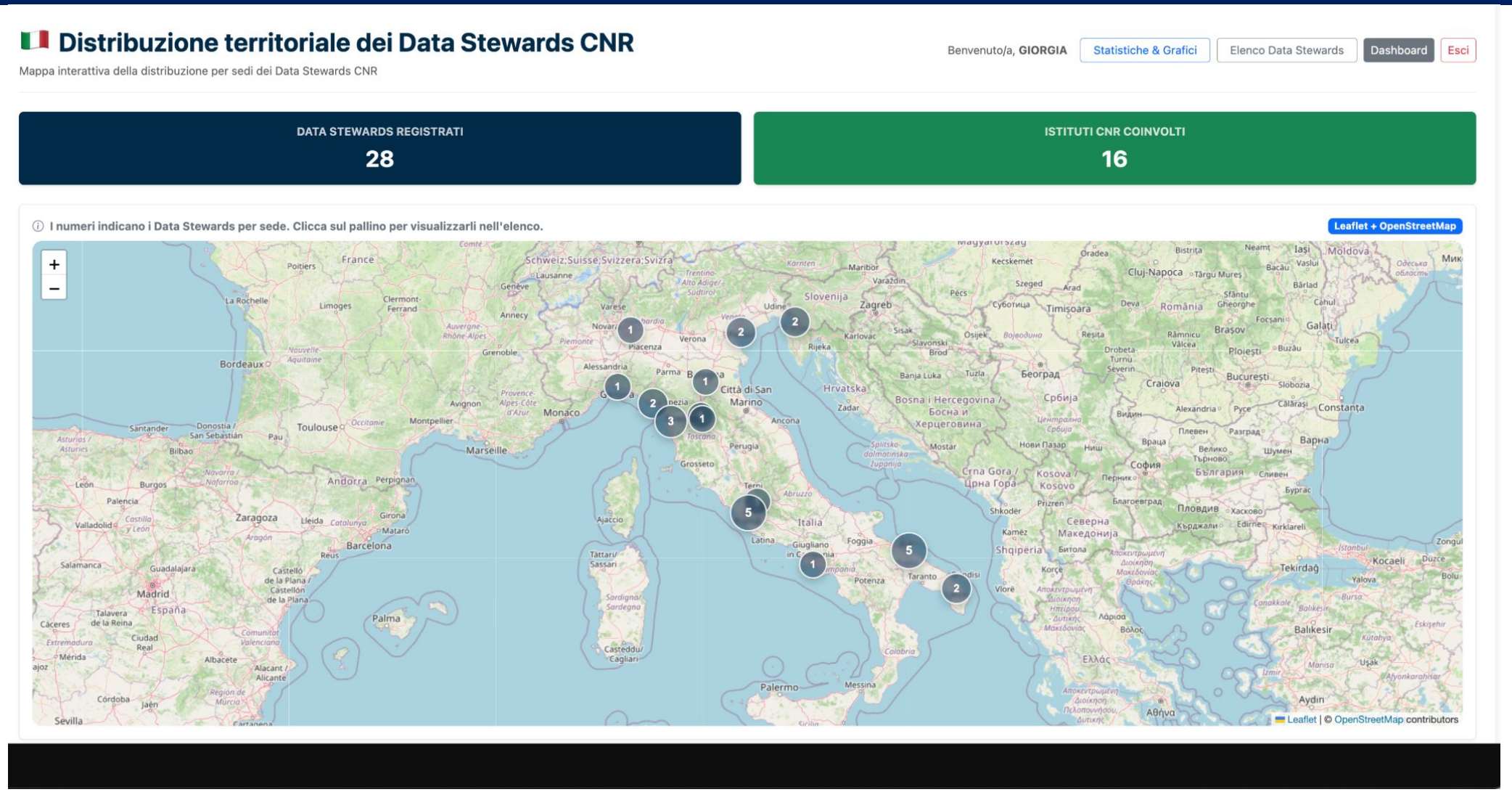


DS@CNR Database: Advanced-Level DS Activities [4/5]



The **keywords** used to **describe competencies** were selected on the basis of the **CODATA RDM Terminology Working Group**.
https://codata.org/initiatives/data-science-and-stewardship/rdm-terminology-wg/

DS@CNR Database: Geographical Distribution [5/5]



Conclusions

- Launched the **first mapping** exercise of Data Stewardship activities at CNR through the establishment of the **DS@CNR Community**, comprising 76 voluntary participants.
- Established a **structured process** for collecting and characterizing **competencies** through a **dedicated survey** and the adoption of **CODATA terminology for classifying Research Data Management activities**.
- Developed a **centralized MongoDB-based database** to manage, browse and update the mapped profiles and competencies.
- Implemented i**nteractive analysis and visualization tools** to explore the distribution of competencies by professional profile, organizational unit, activity area and geographical location.
- **Laid the groundwork for systematic, progressive mapping of RDM competencies across CNR**, aimed at recognizing their value and **supporting the future institutional recognition** of the **Data Steward professional profile**.

Thank You for Your Attention

Would you like to join the DS@CNR Community?

1. **Complete the survey: https://forms.cloud.microsoft/e/72LD8XdWbj**
2. **Register in our database: https://dsdatabase.cnrbiomics.it/**
3. **Contact us!**

Contact: coordinamento-ds@cnr.it

Backup

DS@CNR Database: Technical Details — Authentication

- **OpenID Connect (OIDC): direct handshake with the CNR Biomics Keycloak realm using discovery metadata**
- **Hardening Cookie: SESSION_COOKIE_SECURE, HTTPONLY and SameSite=Lax enabled**
- **@login_required Decorator: granular session-context checks on protected routes.**

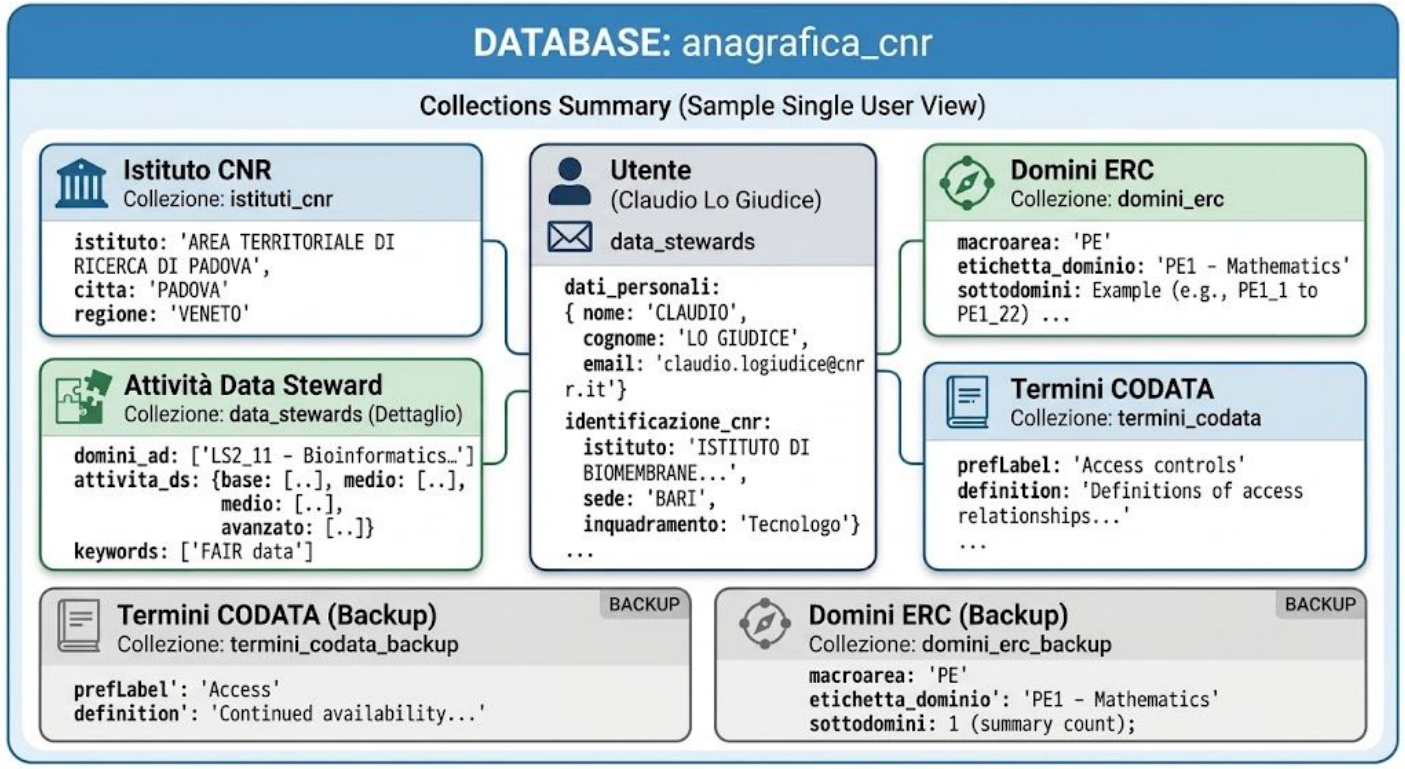
```
// Configurazione Sicurezza Sessione

SESSION_COOKIE_SECURE = True
SESSION_COOKIE_HTTPONLY = True
SESSION_COOKIE_SAMESITE = 'Lax'

// Endpoint Keycloak CNR

server_metadata_url = '.../realms/CNR/...'
```

DS@CNR Database: Technical Details — Data Model




```
{
  "dati_personali": { "nome": "MARIO", "email": "..." },
  "identificazione_cnr": {
    "dipartimento": "DSBA",
    "istituto": "IBIOM",
    "sede": "BARI",
    "inquadramento": "Ricercatore - III liv"
  },
  "identificativi_esterni": { "orcid": "...", "iris": "..." },
  "domini_ad": ["LS2", "LS3"],
  "attivita_ds": { "base": [], "medio": [], "avanzato": [] },
  "consenso_privacy": true
}
```


Idempotent upsert:

Aggiornamento basato su email univoca via Regex Case-Insensitive (**$options: "i"**).

```
profilo_esistente = None
if email_utente:
    profilo_esistente = collezione_data_stewards.find_one({
        "dati_personali.email": {"$regex": f"^{email_utente.strip()}$", "$options": "i"}
    })
```

Tassonomia ERC & CODATA
Struttura ad albero normalizzata memorizzata nelle collezioni ausiliarie.


```
{
  "dati_personali": { "nome": "CLAUDIO", "cognome": "LO GIUDICE",
    "email": "claudio.logiudice@cnr.it" },
  "identificazione_cnr": {
    "dipartimento": "Scienze Bio-Mediche",
    "istituto": "ISTITUTO DI BIOMEMBRANE, BIOENERGETICA E BIOTECNOLOGIE MOLECOLARI",
    "sede": "BARI",
    "inquadramento": "Tecnologo - III liv"
  },
  "identificativi_esterni": { "orcid": "0000-0003-3609-8938", "iris": "" },
  "domini_ad": ["LS2_11 - Bioinformatics and computational biology"],
  "attivita_ds": {
    "attivita_ds_keywords": ["FAIR data"],
    "base": ["Semantic interoperability"],
    "medio": ["Data access protocol"],
    "avanzato": ["Cloud computing", "Data analysis", "Data ingestion", "Data preprocessing", "Integrated access management", "Technical metadata", "Big data"]
  },
  "consenso_privacy": true
}
```