CONFERENZA GARR 2024
Navigare la complessità

The INGV Data Registry, a metadata catalogue to govern Open Science

Mario Locati

Istituto Nazionale di Geofisica e Vulcanologia (INGV)

INGV, a complex data scenario in the field of Earth-Sciences

Approximately 1000 employees, 3 scientific departments (**Earthquakes, Volcanoes, Environment**)
8 main branches (Rome, Milan, Bologna, Pisa, Naples, Grottaminarda, Catania, Palermo)
25 minor branches all around Italy

Data Sources

From the environment
- Instrumental networks
  - terrestrial
  - marine
  - space
- Field surveys
- Laboratories

From society
- Citizen Science

From other organisations

Data Infrastructures

Internal
- Data centers (dynamic and big data)
- Researcher's PCs (static or small data)

External
- Data centers

Data Accessibility

Internally managed
- Data archives
- Web pages
- Scientific Journals

Externally managed
- Data archives
- Scientific Journals

Type
- Basic (files on the Web)
- Advanced (API)

Ownership

Fully owned
- Single author
- Working groups

Mixed ownership
- Public-Public
- Public-Private
- National
- European
- International

Multiple frameworks

Sharing Open Data requires a Research Data Management to govern multiple aspects

Scope and context	Technical frameworks	Legal frameworks	Users
Institutional activities	Data encodings: ▪ Disciplinary ▪ Cross-disciplinary	Licensing	Researchers
Short-term projects	Access APIs: ▪ Disciplinary ▪ Cross-disciplinary	Intellectual Property	Administration
Long-term projects	Identifiers for: ▪ data ▪ organisations ▪ projects ▪ funders	European Data Strategy	Internat. Consortia
	DataCite & INSPIRE Metadata	Spatial Data Directives	Civil Protection
		High Value Datasets	Professionals
		Public-Private contracts	Public

The need of a Data Registry, https://data.ingv.it

Goals

- List and describe all published datasets
- Give credit to authors and funders
- Uniquely identify dataset using DOIs
  (assigned by INGV or another organisation)
- Associate a license (CC BY 4.0)
- Provide a bibliographic data citation
- Link data to papers and software
- Establish a single point of access
- Enforce the institutional Data Policy
  (supervised by a Data Management Office)
- Support the sharing of Open Data

Tools

Publicly accessible

- Web portal
- APIs

Internal use only

- Metadata Editor
- Validation Procedure
- Author's ORCID management
- Projects management
- Organisations management
- DataCite DOI management
- Analysis tools

Data Registry - Workflow

Data Producer	Data Management Office	Scientific Directors	Data Management Office
1. Generate the dataset 2. Publish the data online ▪ INGV infrastructure ▪ external infrastructure, may assign a DOI 3. Create a record using a metadata editor 4. Launch the validation	1. The Metadata Editor send an email 2. Metadata check ▪ Formal check 3. Interact with the Data Producers, if needed 4. Approve or reject the request	1. The Metadata Editor send an email 2. Data check ▪ Scientific check 3. Interact with the Data Producers, if needed 4. Approve or reject the request	1. Assign a DataCite DOI, if not already there 2. Publish the metadata record on the public website data.ingv.it via API (CKAN) 3. (future) push email to a mailing-list

The Metadata Editor is accessible in read-only mode by anyone affiliated to INGV (Google sign-in).

Authors may update their metadata record at anytime, some limitation applied, changes are is checked by the DMO.

Automatic report generation by datasets, author, geographical branch, scientific department, project, funder

After 5 years of use,
the Data Registry has proven to be extremely useful and effective.

The institutional management of Open Science requires dedicated tools to deal with a complex and dynamic scenario.

https://data.ingv.it - ufficiogestionedati@ingv.it