https://doi.org/10.5194/egusphere-egu2020-10057

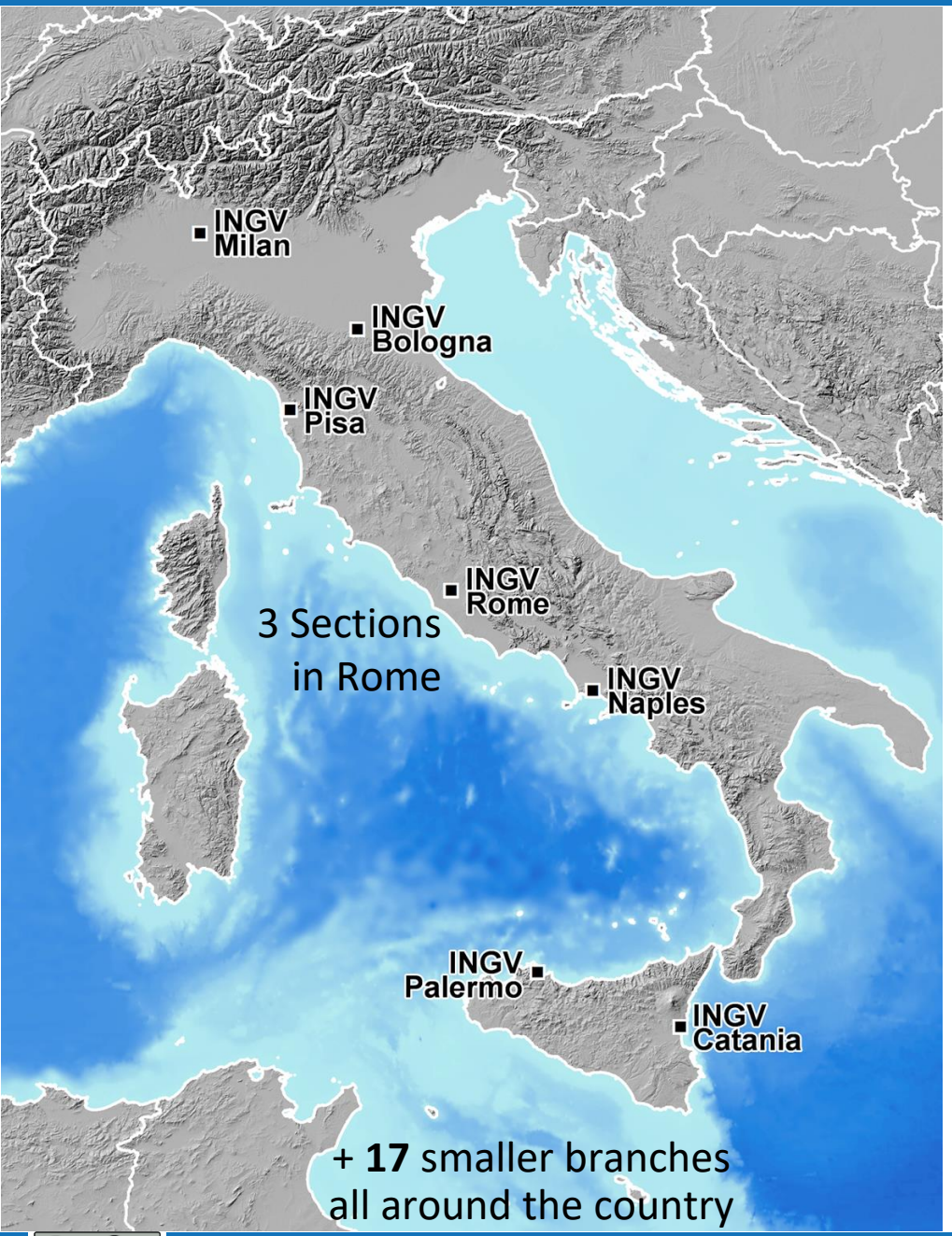
Putting the INGV data policy into practice: considerations after the first-year experience

Mario Locati, F. M. Mele, V. Romano, P. Montalto, V. Lauciani, R. Vallone, G. Puglisi, R. Basili, A. G. Chiodetti, A. Cianchi, M. Drudi, C. Freda, M. Pignone, A. Sangianantoni

EGU General Assembly 2020

CC BY

INGV scientific overview



Established in 2000 by merging pre-existing institutions, INGV is a Public Research Institute supervised by the Italian Ministry of Education, Universities and Research.

3 Departments

- Seismic, GNSS, gravity, magnetic, geochemical networks
- Experimental laboratories for rocks and fluids
- Computational resources
- 3 surveillance centers 24/7 for the seismic, volcanic and tsunami monitoring in cooperation with the Civil Protection
- ~ 800 researchers & technicians

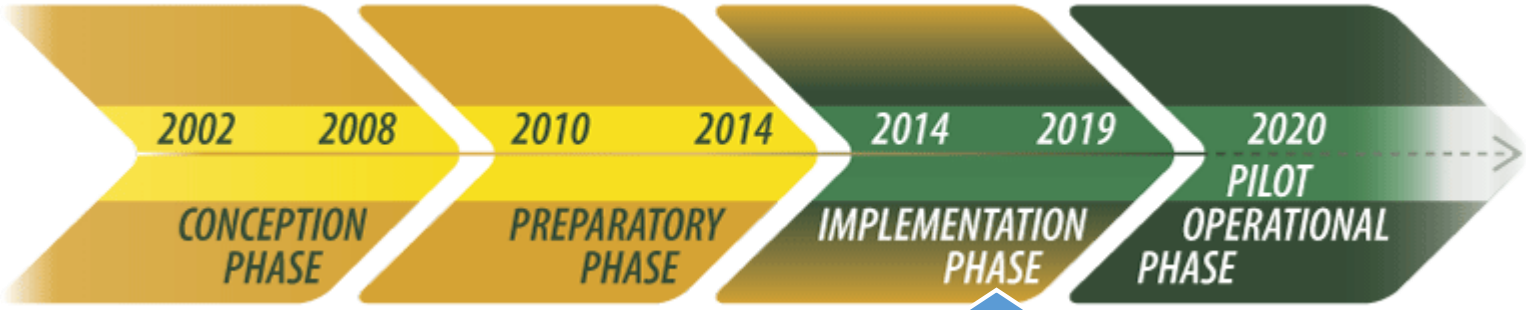
earth prints

Since 2005 https://www.earth-prints.org/ is the INGV Open Access repository used to collect all publications.

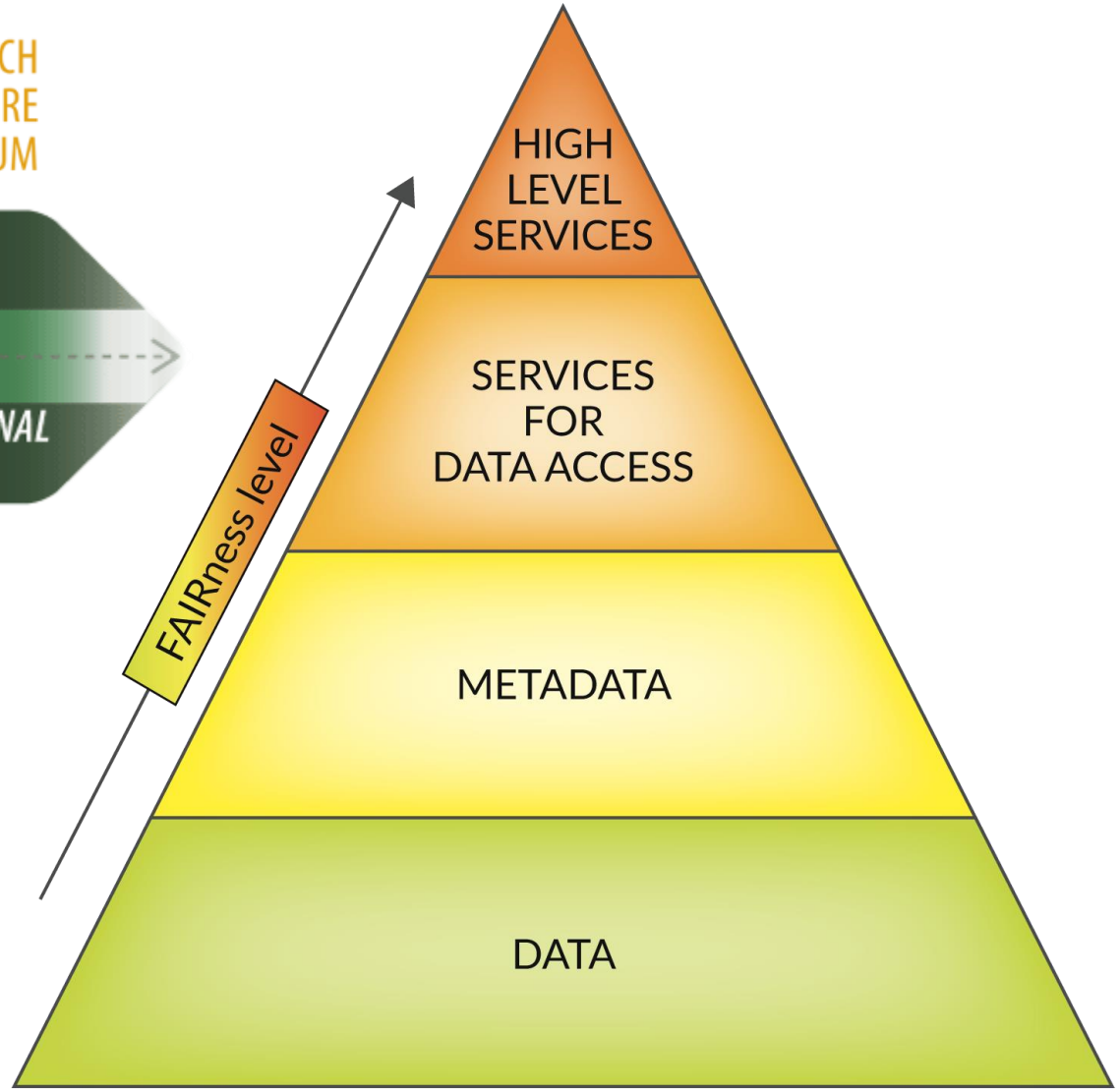
INGV

Extensive re-use of Data in the European Research Area (ERA)

EPOS ERI EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM



2018
EPOS Data Policy



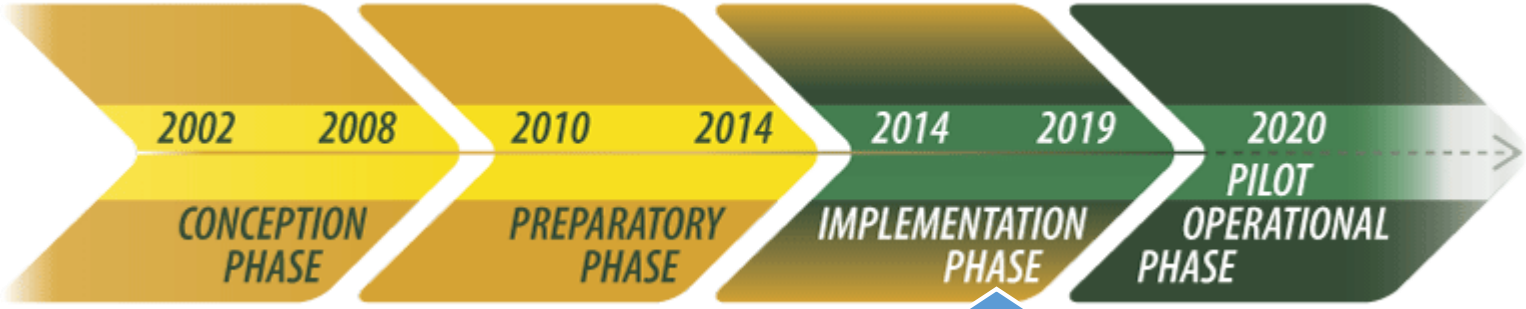
CC BY

INGV

Extensive re-use of Data in the European Research Area (ERA)

EPOS

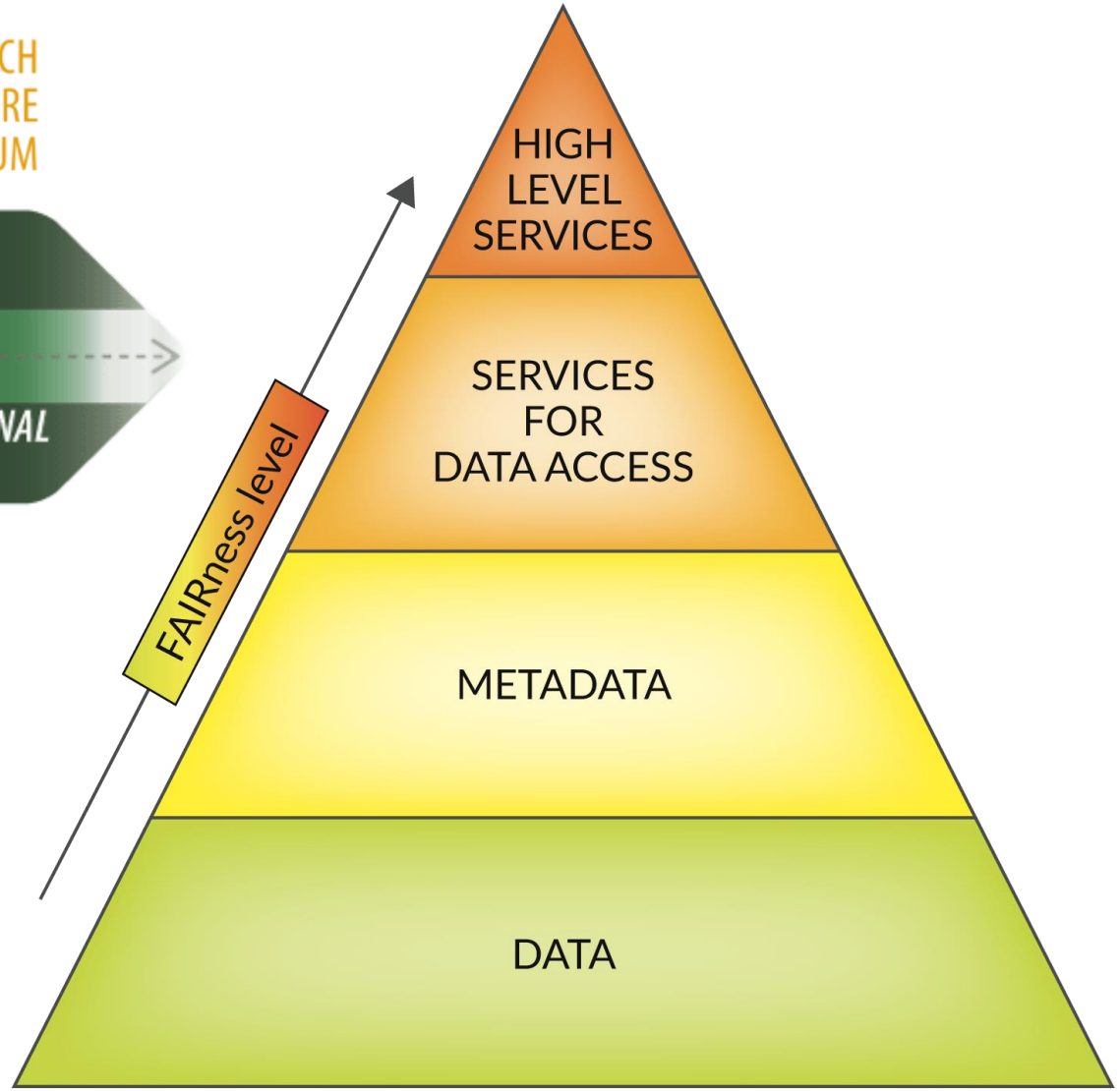
ERI EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM



2018
EPOS Data Policy

EUROPEAN OPEN SCIENCE CLOUD

2015 2020



European legislation on open data and the re-use of public sector information

National Laws | AGID Agenzia per l'Italia Digitale

CC BY

INGV

The INGV Data Policy – https://data.ingv.it/docs/

The INGV Data Policy is made of 3 documents

1. **Data Policy Principles** (2016)
   - Define the area of competence, the legal framework and the main principles inspiring the Data Policy

2. **Open Access Policy on published papers** (2017)
   - Establish a Commission supervising the adoption of the Open Access principles
   - Enforce employees to deposit their works in the Institutional repository (green road) established since 2005, and encourage the publication on the Open Access compliant scientific journal (gold road) run by INGV

3. **Detailed implementation plan for data management** (2018)
   - Provides precise definitions and procedures on how to manage published data
   - Establish the Data Registry and Data Management Office to look after it

CC BY

Data Licensing Scheme

In coordination with the EPOS & EMSO ERICs, INGV selected the **Creative Commons** licenses. The type of license depends on the **level of intellectual contribution** to the data creation.

0 Raw or basic data (e.g. environmental measurements)

1 Data products coming from (nearly) automated procedures

Preferred license

2 Data products resulting from scientific investigations

3 Integrated data products resulting from complex analysis

Authors can suggest another license but must provide a valid reason for it. An evaluation panel approves each suggestion.

Data Management Office

**Part-time involvement*

- Promotes the adoption of Open Science
- Implements and manages the (Meta)Data Registry
- Revises agreements dealing with data in coordination with the legal office

Office Coordinator*

- Plans the activity
- Interacts with scientific and administrative managers
- Interacts with the legal office
- Interacts with the transparency manager

Seismic, Volcanic, Environmental Data Experts*

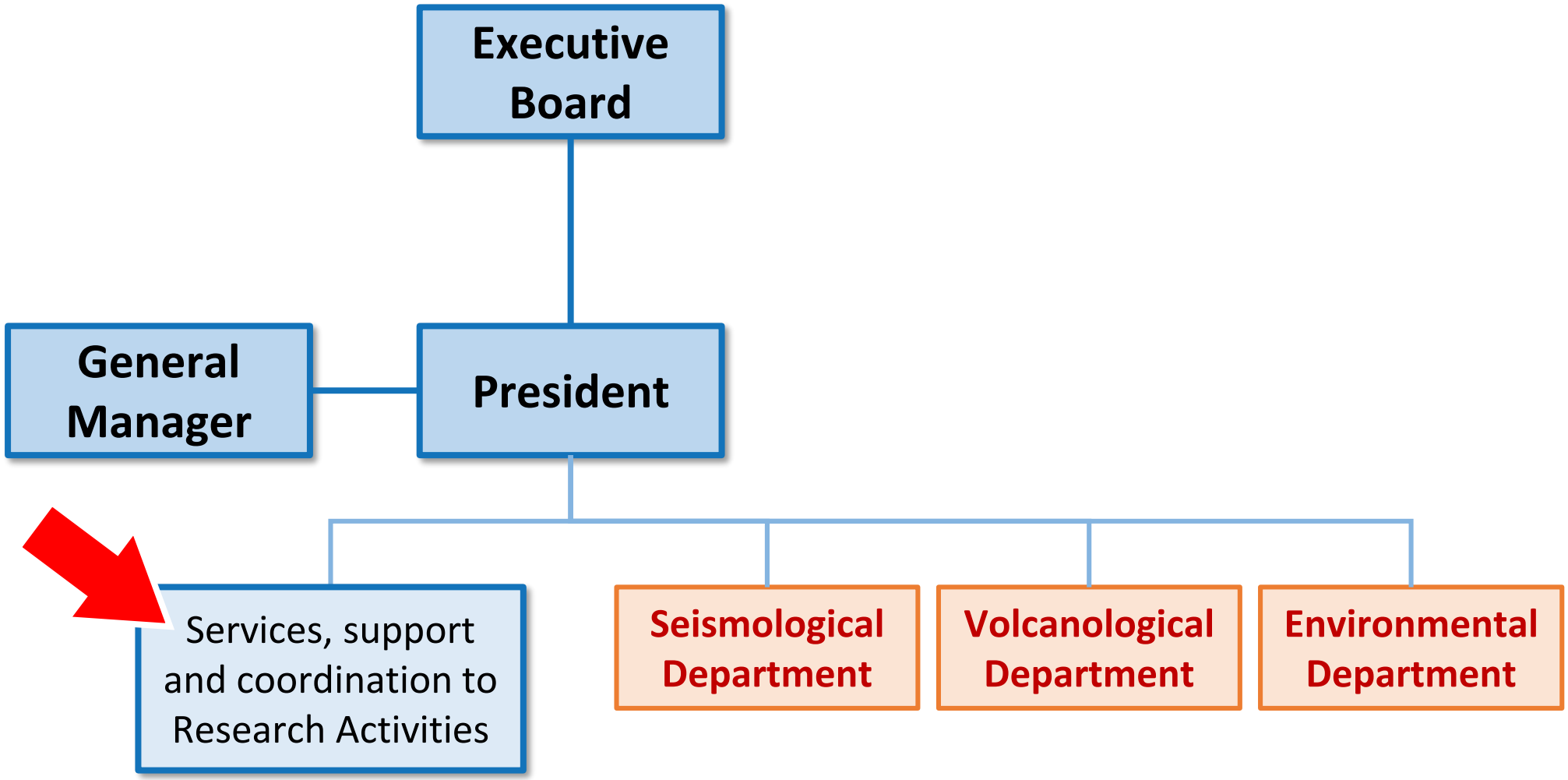
- Deals with disciplinary specific issues
- Supervises the Data Registry scientific content
- Suggests the most suitable standards for harmonizing and optimizing data

IT Experts*

- Supports the coordinator in selecting IT solutions
- Supervises the various existing databases hosting the data
- Interacts with the IT management office

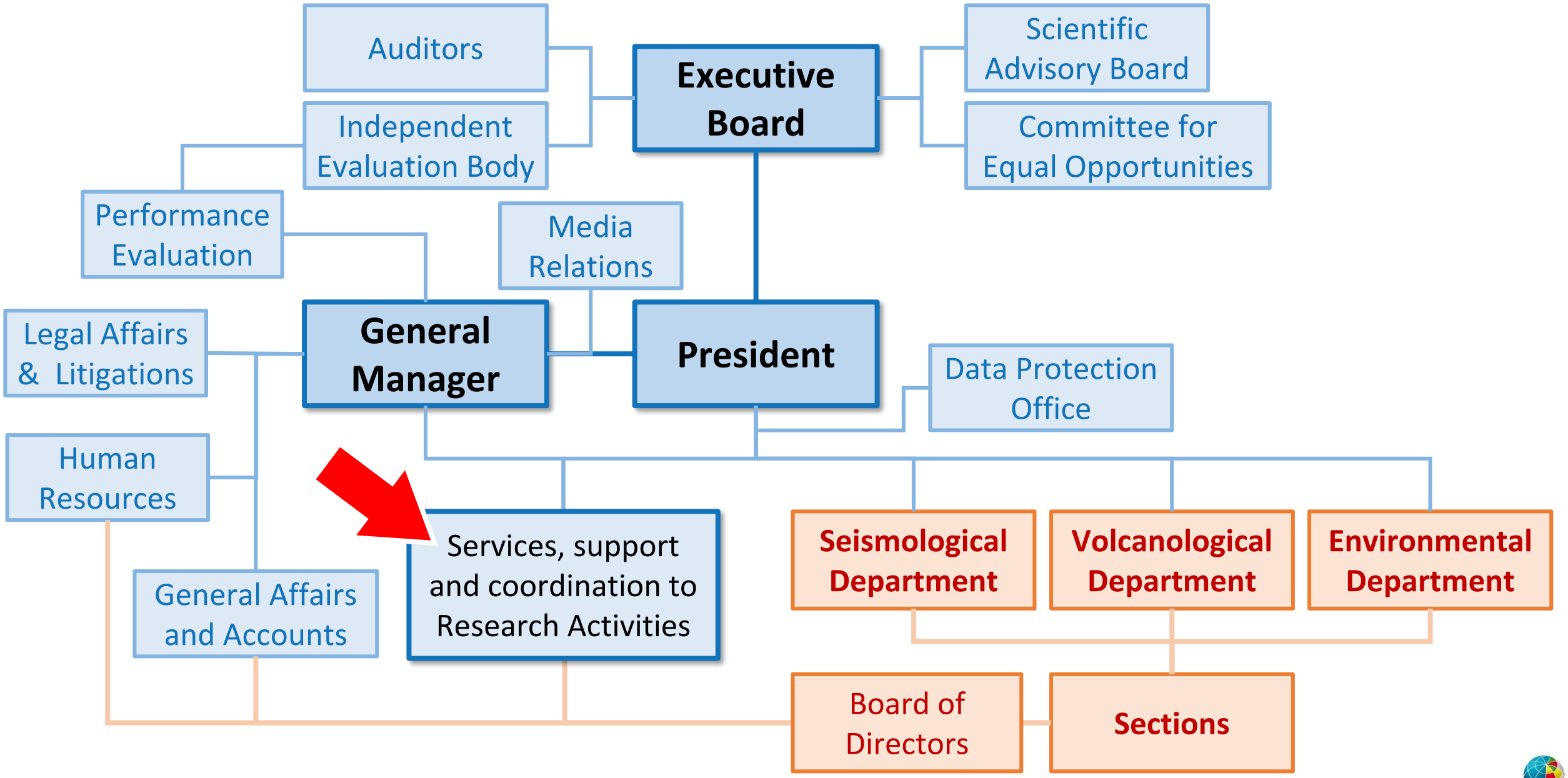
CC BY

Data Management Office and its role in the administration



CC BY

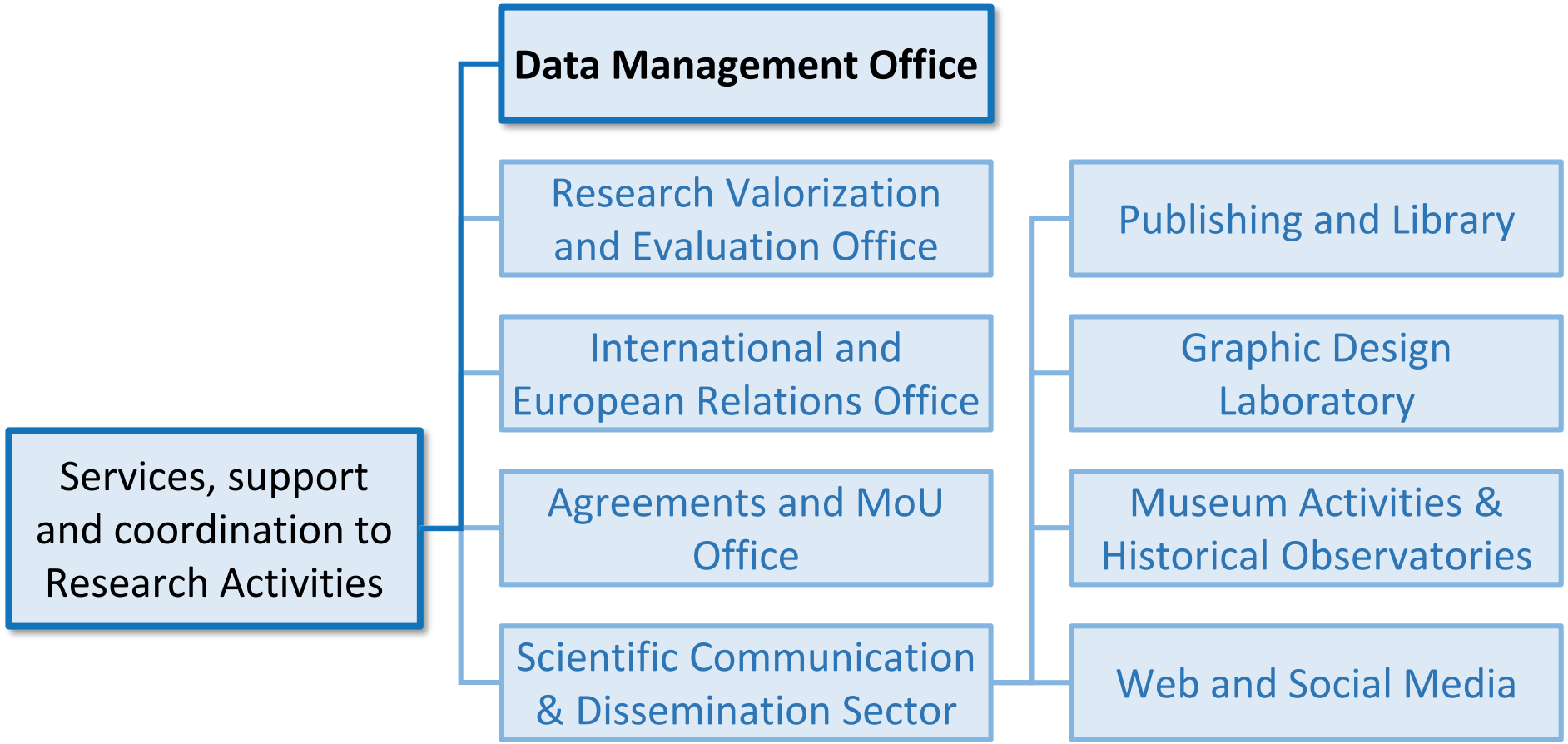
Data Management Office and its role in the administration



CC BY

INGV

Data Management Office and its role in the administration



CC BY

Data Survey

- Lasted 2 months, started in 2016-11-24, ended in 2017-01-23
- Launched by a temporary Working Group in charge of drafting the Data Policy
- Supervised and validated by 9 Section Directors, 3 Departments Directors
- Massive involvement of individuals Researchers
- 275 different types of data identified
- Allowed for getting a better picture of the data generated
- Allowed for establishing an effective classification of available data

CC BY

Data Registry

DOI DataCite registry

Earth-Prints

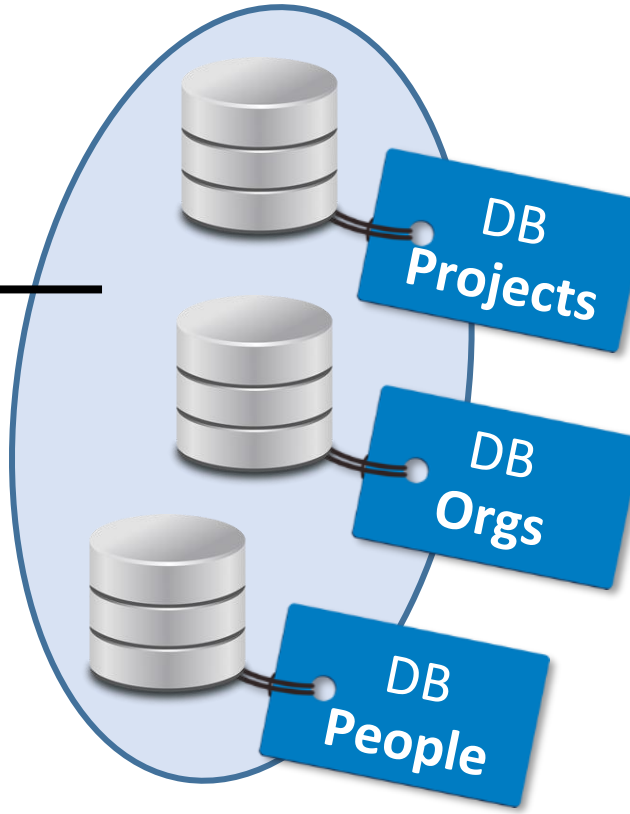
INGV Open Data Portal

Italian **INSPIRE** data portal

INGV (Meta)Data Registry

- metadata about **data access points**
- relations with **other data and papers**
- metadata about **funding projects**
- metadata about **data**
- metadata about **data creators**

Design key rule:
data entry minimisation, entry once, reuse often



metadata

- **download page**
- **web GUI**
- **web services**
- **publications**

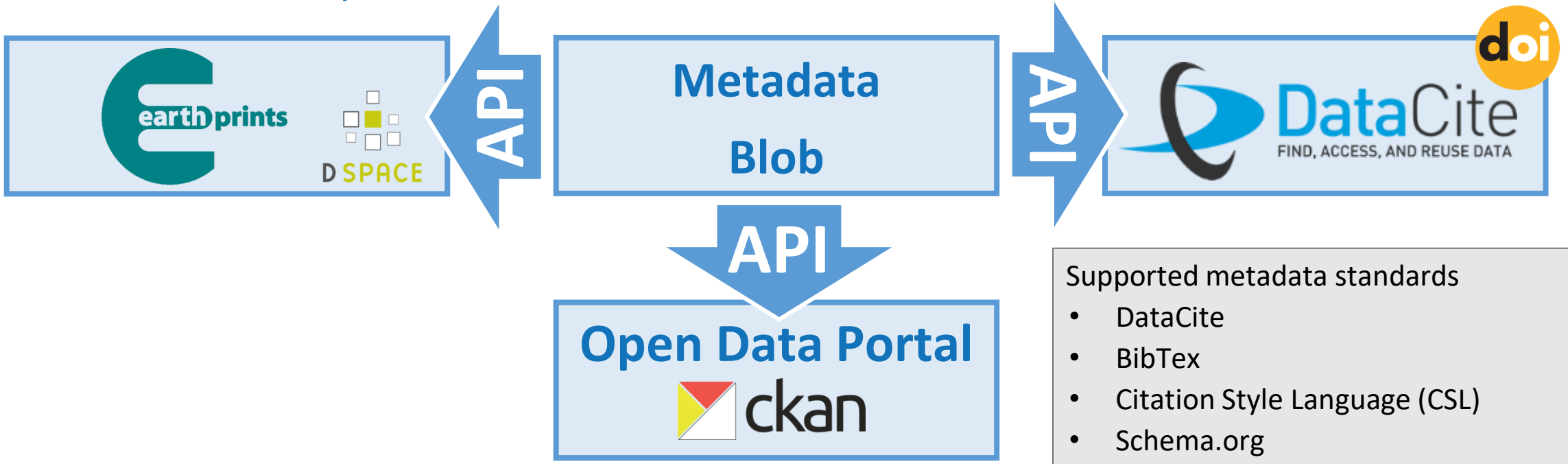
CC BY

INGV

Data Registry as a central hub for publishing metadata on various platforms

National Laws | AGID Agenzia per l'Italia Digitale

↳ Guidelines on web accessibility, metadata standards, data sharing valid for any Public Administrations



Browse content by
- Dataset
- Author

Advanced search engine
CKAN API Support

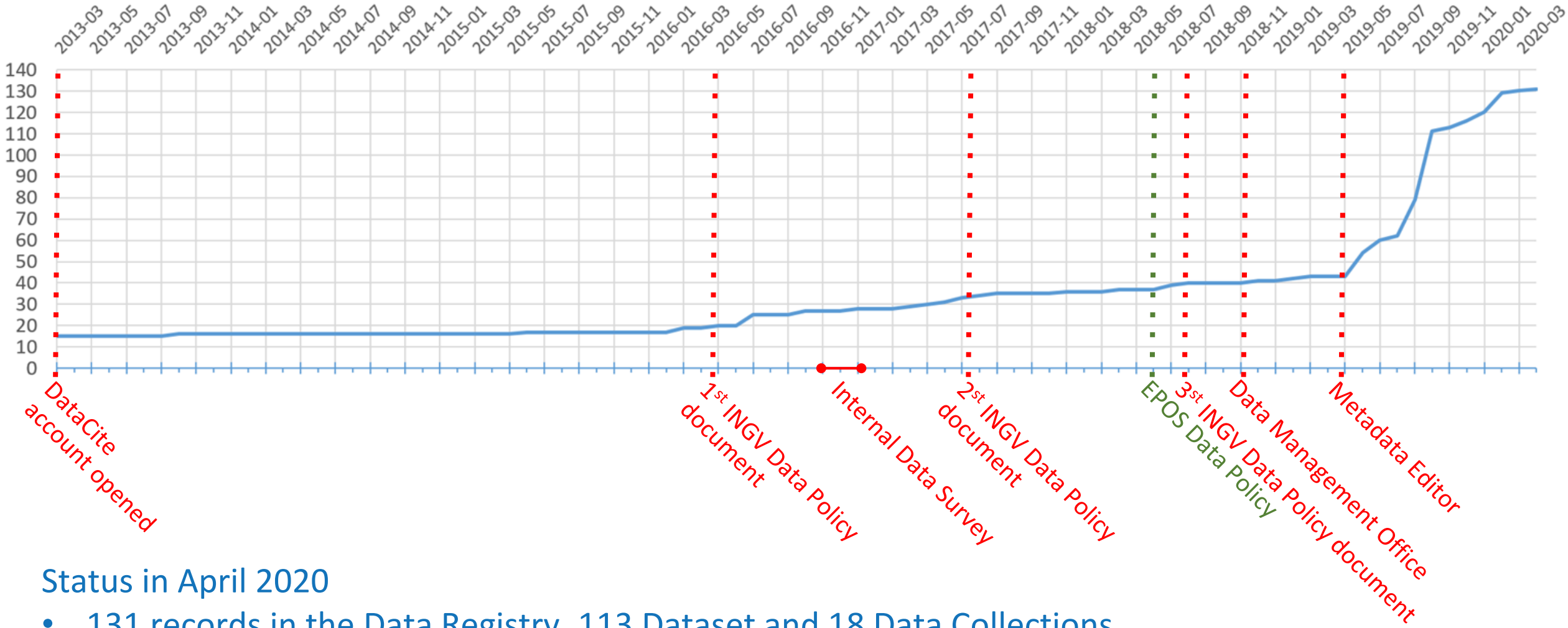
Supported metadata standards
- DataCite
- BibTex
- Citation Style Language (CSL)
- Schema.org
- DCAT-AP

In preparation
- GeoDCAT-AP
- ISO 19115/19139 (INSPIRE)
- NASA DIF

CC BY

Number of minted DOIs

https://search.datacite.org/repositories/crui.ingv



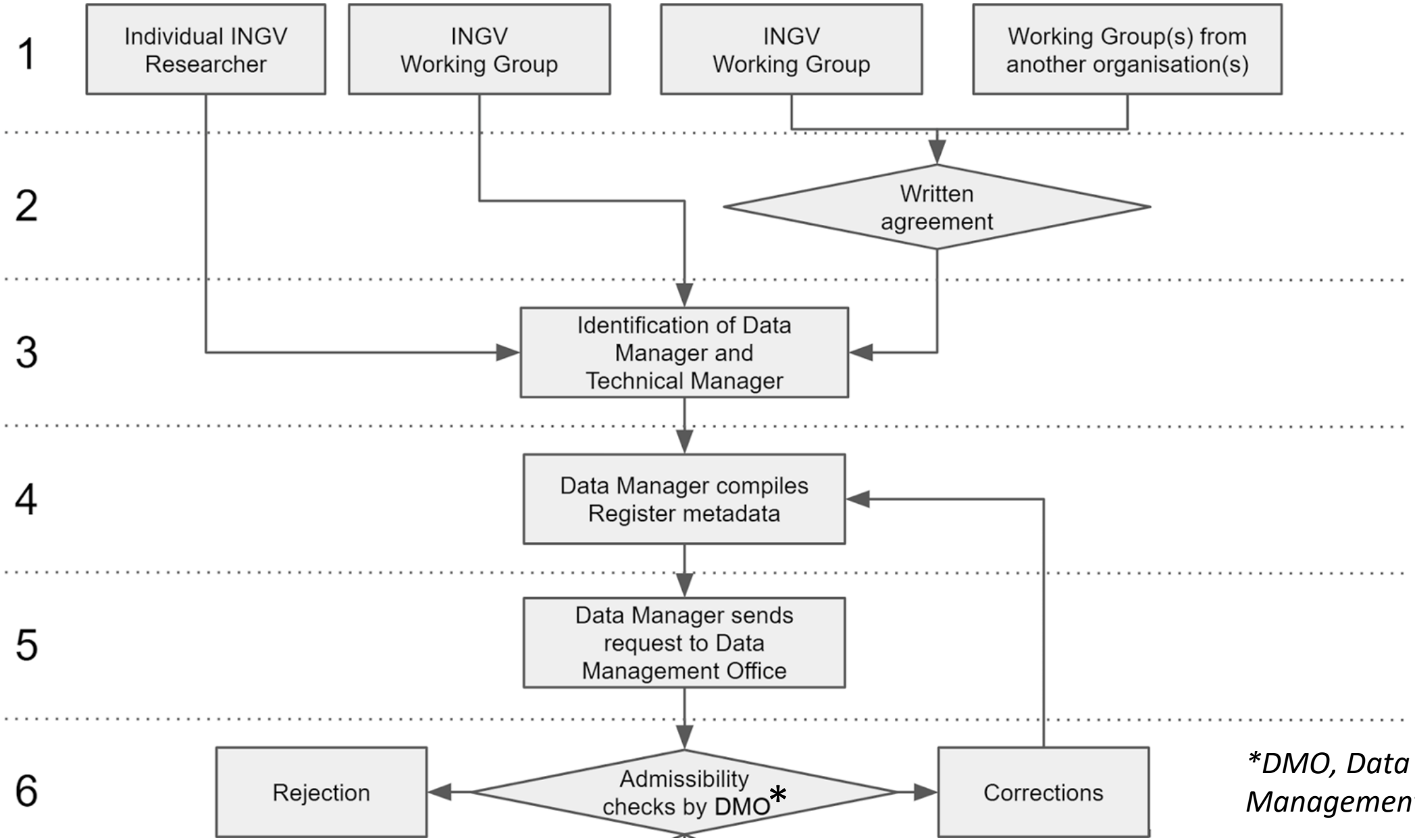
Status in April 2020

- 131 records in the Data Registry, 113 Dataset and 18 Data Collections
- 405 contributing people identified, 71.4% males, 28.6% females, 77% with ORCID
- 66% of people affiliated to INGV, 34% affiliated to 61 other orgs
- 20 related orgs in Italy, 31 in EU, 5 in the Middle East, 2 in North America, 2 in Africa, 1 in Asia

CC BY

INGV

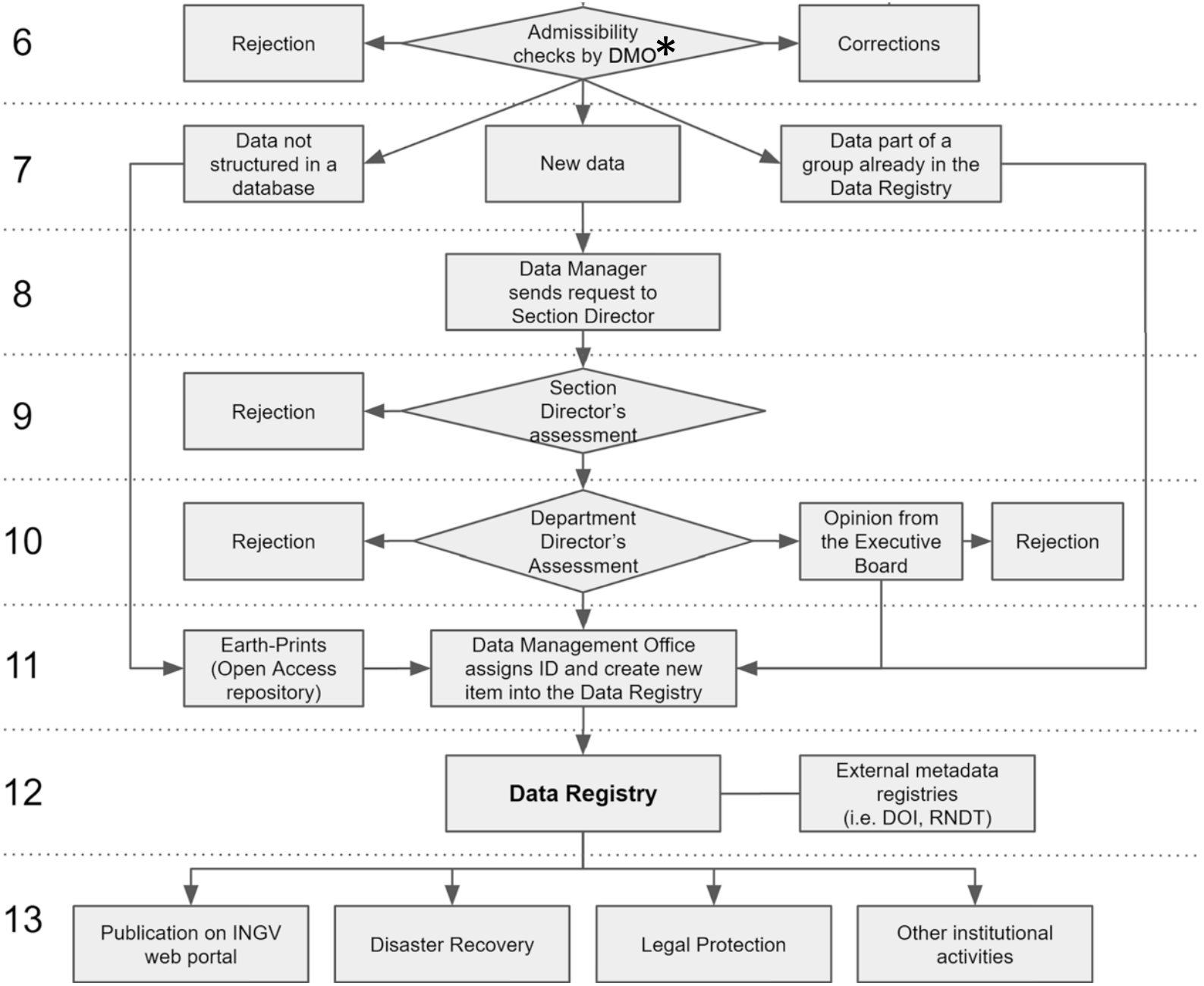
Procedure for entering new data (1 of 2)



**DMO, Data Management Office*

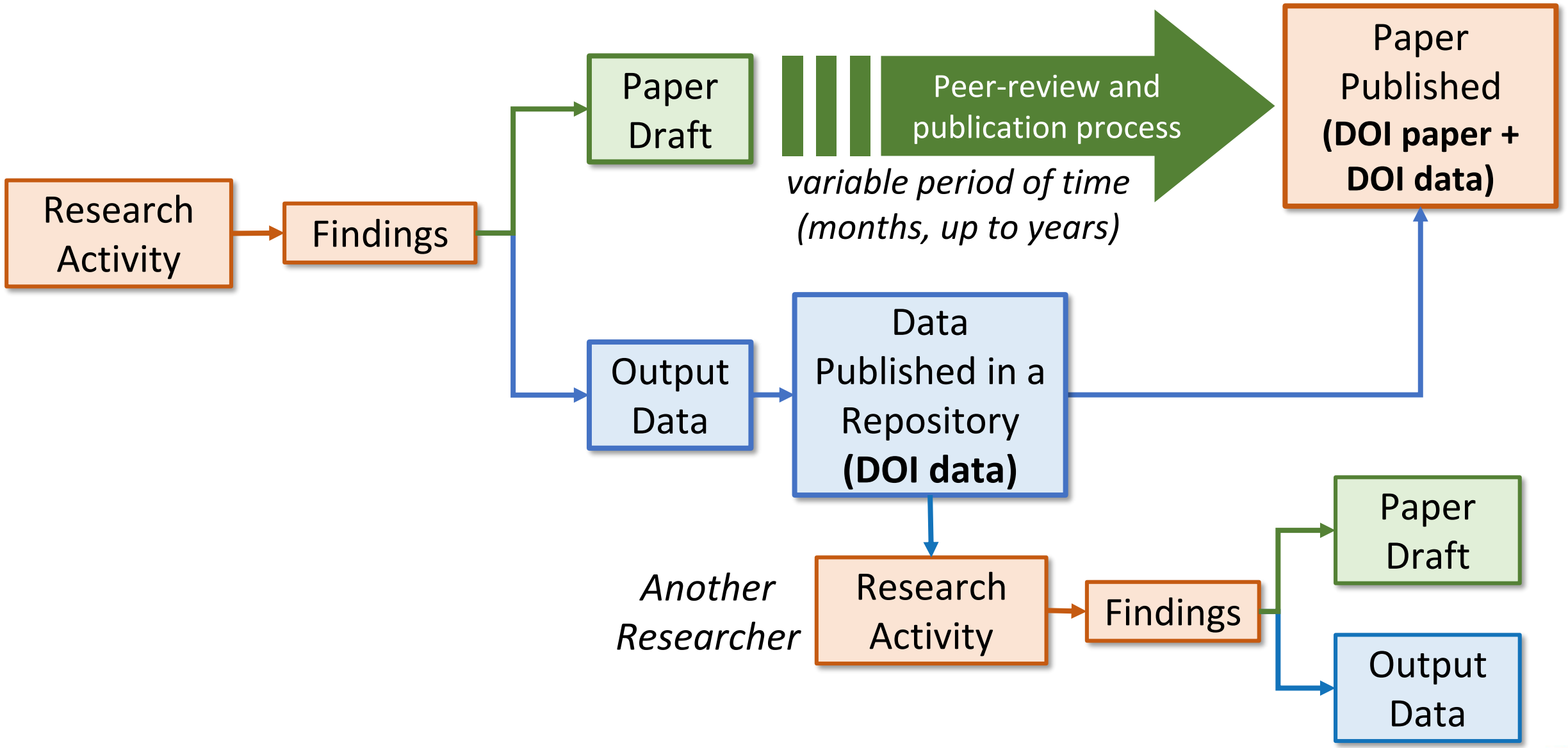
CC BY

Procedure for entering new data (2 of 2)



**DMO, Data Management Office*

Faster data publication will enable Open Science practices



CC BY

INGV

Remarks at the end of the 1st year

Amount of data shared

- Hundreds of datasets emerged from the internal survey, smaller datasets still not traced
- Only a fraction of produced data is publicly accessible, mostly seismological data
- Not -yet- published data is used internally or made accessible to the Civil Protection only

Data encoding and web services

- Delays in publication due to missing data encoding and/or suitable web services standards
- Lack of skills and expertise for investigating OGC standards (i.e. Sensor Web Enablement)

Colleagues response

- Some people still reluctant in openly sharing data
- Overall, data authorship recognition via DOI metadata very well perceived
- Willing to see a tangible recognition during job evaluation processes, inside & outside INGV

Data management

- INGV hosts archives only for specific data, relies on external platforms for data publication
- Building an institutional data repository is under investigation

CC BY

Thank you
for your attention!

datamanagementoffice@ingv.it