



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

**Mario Locati**<sup>1</sup>, Francesco Mariano Mele<sup>2</sup>, Vincenzo Romano<sup>3</sup>, Placido Montalto<sup>4</sup>, Valentino Lauciani<sup>2</sup>, Roberto Vallone<sup>5</sup>, Giuseppe Puglisi<sup>4</sup>, Roberto Basili<sup>5</sup>, Anna Grazia Chiodetti<sup>6</sup>, Antonella Cianchi<sup>7</sup>, Massimiliano Drudi<sup>6,8</sup>, Carmela Freda<sup>5</sup>, Maurizio Pignone<sup>2</sup>, and Agata Sangianantoni<sup>9</sup>

<sup>1</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Milano, Milan, Italy (mario.locati@ingv.it)

<sup>2</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Nazionale Terremoti, Rome, Italy

<sup>3</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Roma 2, Rome, Italy

<sup>4</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Etneo, Catania, Italy

<sup>5</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Roma 1, Rome, Italy

<sup>6</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Bologna, Bologna, Italy

<sup>7</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Amministrazione Centrale, Rome, Italy

<sup>8</sup>CMCC Foundation, Bologna, Italy

<sup>9</sup>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Osservatorio Vesuviano, Naples, Italy

The Istituto Nazionale di Geofisica e Vulcanologia (INGV) has a long tradition of sharing scientific data, well before the Open Science paradigm was conceived. In the last thirty years, a great deal of geophysical data generated by research projects and monitoring activities were published on the Internet, though encoded in multiple formats and made accessible using various technologies.

To organise such a complex scenario, a working group (PoliDat) for implementing an institutional data policy operated from 2015 to 2018. PoliDat published three documents: in 2016, the data policy principles; in 2017, the rules for scientific publications; in 2018, the rules for scientific data management. These documents are available online in Italian, and English (<https://data.ingv.it/docs/>).

According to a preliminary data survey performed between 2016 and 2017, nearly 300 different types of INGV-owned data were identified. In the survey, the compilers were asked to declare all the available scientific data differentiating by the level of intellectual contribution: level 0 identifies raw data generated by fully automated procedures, level 1 identifies data products generated by semi-automated procedures, level 2 is related to data resulting from scientific investigations, and level 3 is associated to integrated data resulting from complex analysis.

A Data Management Office (DMO) was established in November 2018 to put the data policy into practice. DMO first goal was to design and establish a Data Registry aimed to satisfy the extremely differentiated requirements of both internal and external users, either at scientific or managerial levels. The Data Registry is defined as a metadata catalogue, i.e., a container of data descriptions, not the data themselves. In addition, the DMO supports other activities dealing with scientific data,

such as checking contracts, providing advice to the legal office in case of litigations, interacting with the INGV Data Transparency Office, and in more general terms, supporting the adoption of the Open Science principles.

An extensive set of metadata has been identified to accommodate multiple metadata standards. At first, a preliminary set of metadata describing each dataset is compiled by the authors using a web-based interface, then the metadata are validated by the DMO, and finally, a DataCite DOI is minted for each dataset, if not already present. The Data Registry is publicly accessible via a dedicated web portal (<https://data.ingv.it>). A pilot phase aimed to test the Data Registry was carried out in 2019 and involved a limited number of contributors. To this aim, a top-priority data subset was identified according to the relevance of the data within the mission of INGV and the completeness of already available information. The Directors of the Departments of Earthquakes, Volcanoes, and Environment supervised the selection of the data subset.

The pilot phase helped to test and to adjust decisions made and procedures adopted during the planning phase, and allowed us to fine-tune the tools for the data management. During the next year, the Data Registry will enter its production phase and will be open to contributions from all INGV employees.