

PRINCIPLES OF THE INGV DATA POLICY

Data Policy Working Group (PoliDat)

Giuseppe Puglisi (Coordinator)

Roberto Basili

Anna Grazia Chiodetti

Antonella Cianchi

Massimiliano Drudi

Carmela Freda

Mario Locati

Maurizio Pignone

Agata Sangianantoni

13 February 2016

Table of contents

INTRODUCTION	3
1. DEFINITIONS.....	5
DATA	5
METADATA.....	6
DATABASE	6
OPEN ACCESS	6
INTELLECTUAL PROPERTY.....	6
DATA PRODUCER/AUTHOR	6
USERS.....	6
LICENCES	6
SERVICES.....	6
ACCESS	7
PERSISTENT IDENTIFIER.....	7
2. PRINCIPLES	8
OPEN ACCESS	8
IDENTIFICATION OF THE DATA	8
DATA LIFE CYCLE	8
LEGAL MANAGEMENT OF DATA	8
INTEROPERABILITY	9
PRESERVATION	9
ETHICS	9
CATEGORIZATION	9
OWNERSHIP	10
LICENCES	10
CULTURE OF SHARING	10
PROFESSIONAL FIGURES	10
ASSIGNMENT OF PERSISTENT IDENTIFIERS	11
ACCESS RULES	11
MONITORING OF DATA POLICY.....	11

Introduction

The Istituto Nazionale di Geofisica e Vulcanologia (INGV) is a non-commercial public research institution. A legal entity under public law and an autonomous institution, it carries out research activity in the geophysical disciplines in the field of earthquakes, volcanoes, and the environment, with the aim of providing research and responsible innovation while making the research results available to all, in accordance with the principles reported herein, and fostering participation by society.

INGV is responsible for the seismic, volcano, and tsunami surveillance service on the national territory and in the Mediterranean area; it coordinates the activity of regional and local seismic networks; it takes part in European and global studies and surveillance networks; it carries out dissemination activities and promotes communication, information, and training initiatives in schools and companies for the purpose of reducing the risk associated with the phenomena connected to its areas of research; it is a member of the National Civil Protection Service (Servizio Nazionale di Protezione Civile), as well as a competence centre at the Department of Civil Protection (Centro di Competenza del Dipartimento della Protezione Civile), on behalf of which it maintains operative round-the-clock surveillance activities and carries out research projects with dedicated objectives in the area of specific conventions.

INGV has an articulated structure with a strong territorial presence guaranteed by the “Sections”.

In this setting, the requirement to establish a Data Policy arises from the need to:

- Govern the multitude of produced data;
- Set out the inspiring principles upon which to base a shared institutional management of the data;
- Handle the access, use, and reuse thereof.

Open Access is a fundamental principle for INGV, which carries out a major role in society. Consistently with this, INGV, aware of the benefits of Open Access for research in terms of visibility, promotion, and internationalization, signed the *Position Statement sull'Accesso Aperto ai risultati della ricerca scientifica in Italia*¹ (“Position Statement on Open Access to the results of scientific research in Italy”) in May 2013, committing to the effective realization of the principles of Open Access, through the adoption of its own Data Policy.

INGV adopts the Data Policy for allowing open, full, and prompt access to its data while complying with the principles of Open Access established in national and EU regulations, and in agreement with INGV’s institutional purposes, structure, and organization.

The Data Policy establishes guiding documents for:

- INGV’s personnel;
- Departments and Sections;
- National and international bodies that access the produced data in the context of institutional activities;
- Bodies and institutions receiving research products for the purposes of surveillance and

¹ http://istituto.ingv.it/images/Ufficio_Gestione_Dati/docs/20130321_PositionStatementAccessoAperto.pdf

management of natural risks;

- Students and researchers for teaching, scientific, and research purposes;
- Companies;
- Any other party, public or private, or natural or legal body with any interest whatsoever in INGV's data, as defined below.

The set of figures identified above may collectively be referred to as "stakeholders."

The development of data management principles, in compliance with the intellectual property rights and guaranteeing the utmost valorisation and protection of both researchers and results, responds to the public interest of society to:

- Fully access and share the knowledge produced by INGV;
- Contribute to progress;
- Foster the free circulation of ideas.

1. Definitions

The INGV adopts the following definitions in order to attribute univocal meaning to the terms that are used, and to bring the instruments applying the Data Policy in line with the regulations in force.

Data

The term “data” identifies individual items or records of any nature (physical or digital), at any level of processing, and however organized, as well as research products even if unpublished. The term is thus used in this setting in its more general connotation, and refers to publications, the raw data acquired by a sensor, the physical sample of any nature, or a product obtained from any analysis on the data at any level of processing, whether automatic or manual. To bring the definitions in line with the regulations in force, it is useful to bear in mind that the data may be classified as follows:

- Static data: the data, once generated and/or published, are no longer updated;
- Dynamic data: the data may be modified or extended after being published. These data may be further broken down based on the type of variation:
 - data subject to revision, modifiable for the purpose of improving their quality/reliability;
 - continuously produced data, data produced by sensors that generate a time series;
- Spatial data, any geographically localized information²;
- Personal data, any information pertaining to a natural person, legal person, body or association, identified or identifiable, even indirectly, through reference to any other information, including a personal identification number;
- Other types of data, defined as “*all the databases managed for the pursuit of institutional purposes, including those connected with the functioning of the administration*”;³
- Open data, which is to say “usable, reusable, or redistributable freely by anyone, also for commercial purposes, subject at most to the request for attribution and sharing in the same manner”⁴.

² According to the European INSPIRE Directive, these are “any data with a direct or indirect reference to a specific location or geographical area” (EC Directive 2007/2/EC)

³ This is information pertaining, for example, to personnel, financial statements, protocol, document management, etc.)” (Art. 24-quater, paragraph 2, of Legislative Decree no. 90/2014, converted into in Law no. 114/2014).

⁴ The Digital Administration Code (Codice dell'Amministrazione Digitale – CAD, Legislative Decree no. 82 of 07 March 2005.). Art. 68, paragraph 3, defines as open-type digital data those data that:

- Are available in accordance with the terms of a license that permits the use thereof by anyone, also for commercial purposes, in disaggregated format;
- Are accessible via information and communication technologies, including public and private telematic networks, in open formats, are suited to automatic use by computer programs, and are provided with the related metadata;
- Are made available free of charge via information and communication technologies, including public and private telematic networks, or are made available at the marginal costs incurred for reproduction and dissemination thereof.

Metadata

The term “metadata” refers to a structured set of information describing a data. For example, the information may refer to the sensor that provided the data (e.g. position, type, calibration of the sensor), the data acquisition technique (e.g. sampling frequency), to the modes of analysis of the data (e.g. analysis lab, type of model used), or to the context in which the data was generated (e.g. project, ordinary activity).

Database

The term “database” refers to a collection of data, as defined above, that are systematically or methodically arranged and individually accessible via electronic or other means. The protection of databases does not extend to their content, and leaves rights to said content unimpaired⁵.

Open Access

In general, the term “Open Access” refers to the practice of guaranteeing to a final user the access to and reuse of scientific information without additional costs, except for the marginal costs incurred for their reproduction and dissemination.

Intellectual Property

The term “Intellectual Property” indicates a system of the legal protection of intangible assets, in this specific case the data.

Data Producer/Author

The term “data producer or author” is to be understood as the individual research employee or research group that produces the data with means and resources provided by the Institute, including the personnel cost, with both internal (ordinary) and external (for example: from projects, conventions, etc.) funds.

Users

The term “Users” is to be understood as both natural persons and legal bodies, such as research institutions, public or private bodies, associations, companies, etc., that require access to the data for purposes defined by them.

Licences

The Licence is the instrument (for example, contract) defining the rules and conditions to be complied with in order to be able to use the data.

Services

The term “service” is to be understood as any of the following services applicable to a database, such as for example: Data Search, Data Visualization, Transfer, Transformation, Editing, and/or Updating. The individual service available for a database may be:

- Open: the service is freely available and accessible to all, without restrictions;
- Restricted: the service is available, but only under the conditions established or agreed upon by/with the holder of the right to exploit the intellectual property; a particular case is that of data for which a certain type of service (typically the transfer) may be accessed, for a predefined period of time (the “embargo period”).

⁵ D.Lgs. 6 maggio 1999, n. 169 "Attuazione della direttiva 96/9/CE relativa alla tutela giuridica delle banche di dati", GU Serie Generale n.138 del 15-06-1999.

Access

The term “access” refers to a user’s authorization or right to access the data via an infrastructure associated with a database. These accesses may be broken down, in relation to the modes of user identification and registration, as follows:

- Anonymous/Guest: access to the service(s) takes place with no identification or accreditation procedure;
- Registered/Standard: access to the service(s) requires an identification by means of accreditation procedures which may be automated;
- Authorized: access to the service(s) requires specific authorization by the database manager.

For the purposes of IT security, the access procedures have three functions: 1) user Identification, 2) Authentication of credentials, and 3) Authorization to use some services.

Persistent Identifier

The Persistent Identifier is an alphanumeric code whose structure is defined within objects management systems (data, books, persons, physical samples, etc.), some of which follow open and internationally shared standards (Open Standard). The adjective “persistent” refers to the characteristic of guaranteeing that the association between the associated object’s identification code and its position on the Web is maintained over the long term. This means not only that the identifier management, but also access to the associated object, are maintained operational over time. A data may be associated with one or more types of Persistent Identifiers.

2. Principles

INGV, aware of its aim to achieve an open, collaborative, and international research environment based on reciprocity, adopts the following Principles for the management of the Data produced in the context of research activities financed with public funds:

Open Access

INGV adopts the principle of Open Access to research data produced in the context of research activities financed with public funds, in order to:

- permit free, open, full, and prompt access to its data;
- improve the quality of the data, and the pertinence, acceptability, and sustainability of the results;
- reduce the possibility of duplication of research activities;
- accelerate scientific progress;
- give stakeholders a way to interact in the research cycle, by integrating society's expectations, needs, interests, and values.

Identification of the Data

INGV is committed to making the produced data publicly accessible, usable, and reusable, in the context of research activities financed with public funds, using special electronic digital infrastructure(s).

The data will be identifiable and linked to the corresponding metadata to provide all the elements necessary for the proper use thereof.

Data Life Cycle

INGV aims to promote and develop methodologies that make enable an arrangement common to all the phases in the data life cycle:

- collection / acquisition;
- processing / analysis;
- archiving / preservation;
- access;
- reuse.

Legal Management of Data

INGV is committed to applying and adopting the complex and stratified national and EU regulations in force regarding the availability of the data of public administrations, data access services, the regulation of intellectual property, the modes of the exploitability of the data, and the criteria for managing open data⁶.

⁶It appears difficult to include the multitude of legislative interventions in the matter. Therefore, for the purpose of systematizing and making more immediately perceptible the state of the regulations in force, a summary table was prepared, including a regulatory framework existing on the international, European, and national level on the date of the drawing-up hereof (Attachment 1).

Interoperability

INGV is committed to promoting compliance with national and EU standards of semantic interoperability among databases and to adopting open formats and standards for the coding and transfer of the data and of the data catalogues, understood as public, comprehensively documented, and neutral with respect to the technological instruments necessary for exploiting these data.

Preservation

INGV promotes the long-term preservation of the data in digital format, in order to respond to the public interest (referred to as “digital-first”).

To achieve this principle, it is committed to implementing a related and progressive associated financial planning.

Ethics

INGV is committed to adopting ethical behaviour to ensure responsible management of the data, making the research results available to the stakeholders and fostering the participation of society. INGV is also committed to adopting the measures necessary to mitigate the risk of abuse of the data, or of inappropriate use of its data and results.

Categorization

In order to proceed with the definition of the principles of appropriate data management, INGV makes the following categorization:

Publications: The term “Publications” is to be understood as all the research products published in magazines, books, or series of books, following the respective editorial lines.

Research Data: INGV categorizes the research data in four Levels distinguished based on their degree of processing:

- Level 0: raw data, or basic data, acquired automatically or manually, with no level of processing, excluding an automatic-type validation (examples: waveforms, GPS data, uncalibrated images, rock samples, etc.);
- Level 1: data products obtained from automatic or semiautomatic procedures (examples: localization, magnitude, focal mechanisms of earthquakes, shakemaps, historic series of the amplitude of volcanic tremor, and of movement of GPS stations, etc.);
- Level 2: data products obtained from the research activity and at any rate based on non-automatic procedures (examples: crust models, strain maps, source models of earthquakes and of deformations, results of geophysical campaigns, etc.);
- Level 3: integrated data products obtained from complex analyses that integrate several Level 2 products, or from analyses that integrate Level 1 or 2 products of different types and/or originating from different communities (examples: hazard maps, catalogues of active faults, volcanic activity reports, etc.).

For the needs of clarity and the valorisation of professional figures, INGV is committed to instituting a Research Data Registry classifying the data within the Levels defined above, associating each with the attributes necessary for managing ownership, Licenses, Persistent Identifiers, access rules, and the rules for accessing the necessary links to the source of the data.

Ownership

To balance the rights and the complex of opposing interests, INGV accommodates and adopts the Open Access paradigm aimed at fostering public access to information and data sharing, where necessary compromising individual needs in favour of the public interest.

For Level 0 and 1 data, considering the restricted intellectual property thereto, INGV assumes full ownership and therefore the pro tempore Legal Representative manages its complex of rights such as the attribution of Licences, the assignment of Persistent Identifiers, and the signing of agreements for the management and use thereof.

For Level 2 and 3 data, INGV is committed to adopting a Regulation for attributing authorship of a specific data to an INGV individual or working group employee.

The ownership of the publications is established based on the regulations in force and on those of the journals in which the work is published. INGV recognizes, as publishing models, the proprietary one adopted by most scientific publications, in which the author grants the publisher the right to publish the original work, also transferring the right to its reuse, and the Open Access model ("Green road," "Gold road") according to which the author maintains the rights to his or her work for the recognition of ownership, for a greater dissemination of the results and to allow its reuse.

Licences

INGV promotes the adoption of Creative Commons Licences (see Appendix 2) for the management of the rights to its own data. For Level 0 and 1 data, the party qualified to grant the licence is the Institute, by way of its Legal Representative.

Culture of Sharing

INGV is committed to supporting researchers who adhere to a culture of sharing of the results of their research activities, encouraging and promoting the institution of new measurement criteria and new indicators.

Professional figures

For the achievement and effective implementation of data management principles, INGV is committed to the allocation of professional resources, defining the necessary figures and the corresponding responsibilities⁷.

9

⁷ The Agency for Digital Italy (Agenzia Digitale per l'Italia, AgID) Guidelines (2014) identify the following professional figures for an appropriate data management:

- Open Data Manager, with abilities to manage both the technical side and the coordination for the various activities conducted within the group, and to interface with such outside figures as the Transparency Manager (Legislative Decree no. 33 of 14 March 2013);
- Database Manager, the executive figure responsible for populating the portal for accessing the Body's data;
- Database technical contact, the operative arm of the database Manager who keeps the technological infrastructure operative;
- Database thematic contact, the scientific contact who examines the content and interfaces with the authors of the data;
- Statistical Office, which manages the data's display and monitors their use, while also promoting the inclusion of new data, where necessary;

Assignment of Persistent Identifiers

INGV is committed to adopting persistent identifiers for the purposes of achieving:

- Univocal identification of the data, regardless of its position;
- The traceability of the data, both inside and outside INGV;
- Durability and reliability over time, with the related burden of sustainability;
- Certifying the data's authorship and consequently their degree of authoritativeness;
- A mapping of the relationships both among own data and with data generated externally;
- Facilitating the protection of investments.

In order to effect an informed management of the attribution of the persistent identifiers, INGV will have a Regulation for this purpose, containing operative instructions on the applicant, assignment of licences, categorization, and all else needed for effective attribution.

Access Rules

INGV is committed to defining, for each category and type of data, the procedures for access by users who will have to adhere to specific access rules and conditions preliminarily to Access on the digital/electronic infrastructure(s).

The Access Rules shall be established and defined after the development of the electronic and digital infrastructure(s) that will manage access to the respective data.

Monitoring of Data Policy

INGV is committed to monitoring and updating the Principles by means of a Working Group for this purpose, which will have to bring the INGV Data Policy in line with the ongoing development of the regulations in force and the technical/scientific framework of reference on an internal, national, and international setting, define the procedures for the proper management of the data, guarantee their application, and promote the training of personnel for the purposes of applying the Data Policy.

-
- Juridical/administrative office, which deals with the licences, legal notes to be associated with the data, problems related to intellectual property, privacy and the management of personal data, restrictions on use and access;
 - Communication team, with the professional figures for managing institutional communication, both outside the Body and within it.