Geodetic Data Publication and Citation Strategies using DOIs with Examples from GFZ Data Services

Kirsten Elger

GFZ Data Services
GFZ German Research Centre for Geosciences, Potsdam, Germany
What is a DOI?

• Digital Object Identifier

• A globally **unique** and **permanent** identifier for digital objects (link)

• With **metadata** and **licence**

• **Persistent** = long term data access guaranteed by the publisher

**DOI-referenced datasets:**

• are directly accessible via the DOI link

• link to the exact data used for research results → reproducibility

• are citable in scholarly literature → tracking, → credit for researchers and institutions

• DOI metadata increases data discoverability (via catalogues)
A brief history of DOIs

- 2000  first DOI system for online articles (implemented by Crossref)
- 2004  **first DOIs for data minted** (DKRZ, GFZ, Pangaea)
- 2007  Foundation of DataCite (DOI registration agency for data, software and grey literature)

**STD-DOI (2003-2008)**
Publication and Citation of Scientific Primary Data

**DFG**-funded project developing the application of the DOI concept for data
Recent agreements between publishers, repositories, funders, ... 2015-today

- Data availability statement required for articles
- Citation of data in reference lists allowed by journals
- Recommendation to use domain repositories
- No data supplements anymore (→ data publication via repositories)

ENABLING FAIR DATA COMMITMENT STATEMENT IN THE EARTH, SPACE, AND ENVIRONMENTAL SCIENCES

2015

COPDESS

2018
GFZ Data Services – domain repository

Profile
• Domain repository for the Geosciences
• DOIs for Data and Software
• Data: real-time data streams (observational data), tables, maps, model data, ...
• Data reports with DOI, internal review

Technical „Highlights“
• International metadata standards for discovery metadata (human & machine readable)
• Controlled Vocabularies for „rich“ metadata
• PIDs (DOI. ORCID, Fundref, IGSN)
• Open Licences for data and software
• OAI-PMH interface
• schema.org → Google Dataset Search

http://dataservices.gfz-potsdam.de
DOI minting strategies – easy for static data

- Static data don’t change
- New DOI versions for updates
Observatories and monitoring networks @ GFZ

But what about dynamic data?

**GNSS data**
GFZ is operating an analysis center of the International GNSS Service (IGS). Data from a globally distributed network of more than 200 stations serve as a basis for determining the spatial and temporal variation of orbit parameters, rotation parameters, and polar motion parameters. The data are processed to provide information on the Earth's rotation, polar motion, and motion of the Earth's center of mass. Presently the navigation service is provided.</p>

**Seismological data**
The global seismic network GEONET offers seismic data and earthquake data. It includes more than 1000 sensors and recording devices. The center records and distributes all data in real time. The catalogue services provides data of all earthquake events which have been recorded by GEONET sensors. The data is freely accessible on the Internet.

**Ground deformation monitoring**
The German Geodetic Monitoring Network (GGMN) provides monitoring data of ground deformation at several sites in Germany. The data is used for geodynamic research. The centers of the network are accessible on the Internet.

**GPS**

**Satellite data**
GFZ analyses data of different satellite missions, which describe the geophysical parameters of the Earth. Data from the CHAMP and GRACE missions are long-term archived at GFZ. The center provides data sets together with an analysis service in particular for CHAMP data.

**Earth's gravity field data**
The International Centre for Global Earth Models (ICGEM) makes all global earth models available, which are provided as sets of spherical harmonic coefficients. Some data are available to the public. This data set includes permanent geodetic data and historical data. The spherical harmonic coefficients are available in a standardised self-explanatory format. The models can be downloaded from the ICGEM web site.

**Seismology**
The planetary three-hour-range Kp index was introduced by Bartels in 1949 and is derived from the data of 13 magnetic observatories. It is designed to measure solar particle radiation by its magnetic effects. The geomagnetic Ap, Ap, Cp and Kp indices are calculated for each day.

**Geomagnetics**
The GIPP Experiment and Data Archive is the platform for long-term archiving of geophysical observations. The GIPP national data archive contains the German geomagnetic and geoelectrical data. The "Geophysical Instrument-Pool Potsdam (GIPP)" contains data of controlled-source seismic and magnetotelluric field measurements.
A special note regarding citation of dynamic datasets:

For datasets that are continuously and rapidly updated, there are special challenges both in citation and preservation. For citation, three approaches are possible:

a) Cite a specific time slice (the set of updates to the dataset made during a particular period of time);

b) Cite a specific snap shot (a copy of the entire dataset made at a specific time);

c) Cite the continuously updated dataset, but add an Access Date and Time to the citation.

Note that a “time slice” and “snap shot” are versions of the dataset and require unique identifiers. The third option is controversial, because it necessarily means that following the citation does not result in observation of the resource as cited.
What to keep in mind when minting DOIs for dynamic data?

Guarantee reproducibility of published results

- **Data that are published must not change** → reprocessing of online data require a new version
- **Save a copy of all DOI versions** for later requests (not necessarily online)

Granularity

- **DOI should provide reasonable citations** (who would like to cite 100 DOIs for GNSS stations of a network in a paper?).
- The composition of a data publication „package“ should fit to the **usage of data**
  → DOIs for attribution
Why Seismic Networks Need Digital Object Identifiers

In a move to give credit where it's due, the International Federation of Digital Seismograph Networks will link digital object identifiers to data from seismic networks and project deployments.

We hope that assigning seismic data networks a universal and easily cited digital identity will help bring data providers the recognition they deserve.

- 128 DOIs for seismic networks, 10 data centers, 8 countries (13 July 2019)
- Standardised metadata: FDSN Recommendations for seismic network DOIs (https://doi.org/10.7914/d11596)

DOIs for Dynamic Data in geodesy

DOIs for GPS/GNSS Datasets and Networks

Station/Observatory DOIs for the International Geodynamics and Earth Tide Service (IGETS, superconducting gravimeter data) + Data report with DOI

DOIs for ICGEM temporal models

GRACE/GRACE-FO data and products (time series)
Dois for geodetic data – what is already there?

- **DOIs for GNSS Data:** many troposphere products
- **GPS Networks:** mostly UNAVCO, campaign data (static), but: cGPS data for RING, IPOC
- **IVS, Doris:** contributions to ITRF
- **IGETS:** growing time series
- **Many supplements to papers, but also Observatory data**
Questions for the DOI Working Group

• What do we want to assign DOIs with? Data, Products, Networks, Services?
• What are the best practices for DOI-referenced geodetic data already in use?
• Do we need DOIs for all or would a combination of DOIs for attribution and other PIDs (e.g. Handles for identification of files) be more appropriate?
• What about „technical“ metadata (station logs) – how can we best relate the data with metadata? Shall we assign PIDs to the station logs?
• PIDs allow machine-readable cross-referencing (provenance)
• What about PIDs for instruments?