Digital Object Identifiers for the IVS

Glenda Coetzer[1,2], Yu Takagi[3], Kirsten Elger[4]

Abstract One of the goals of the International VLBI Service for Geodesy and Astrometry (IVS) is to provide data and products to support geodetic, geophysical and astrometric research, and operational activities (IVS, 2022). The IVS is committed to supporting scientific discovery through good data management. To enhance data visibility and sharing, IVS data and products need to adhere to the FAIR (Findable, Accessible, Interoperable and Reusable) data principles. In support of FAIR data, the IVS Directing Board agreed on the use of persistent identifiers, *i.e.*, Digital Object Identifiers (DOIs), for permanently identifying its data and products. We provide feedback of an exploratory study that is being conducted to establish best practices for attributing DOIs to IVS data and products.

Keywords DOI, research data management, research data repository

1 Introduction

Research institutions and scientists are evaluated on the scientific output (*i.e.* scholarly literature and citation numbers) they produce (Bordelon, Grothkopf, and Meakins, 2018). Compared to other scientific disciplines, geodesy researchers appear to be producing less "countable scientific output". This is because geodesy researchers are much more involved in operational aspects of institutions, data creation, and provision (Elger *et al*., 2020). It is imperative to establish structured and well-documented mechanisms for geodesy data, products, software, equipment/instruments, stations, and networks, which can assure discovery, retrieval, and citation of data used in scientific publication, and to give recognition to individuals, institutions and funders for the creation and storage of data (Bordelon, Grothkopf, and Meakins, 2018; Elger *et al*., 2020).

1. South African Radio Astronomy Observatory
2. University of Pretoria, South Africa
3. Geospatial Information Authority of Japan
4. GFZ German Research Centre for Geosciences

2 Digital Object Identifiers

Four years after the implementation of Digital Object Identifiers (DOIs) for unambiguously identifying and linking to online articles, the first DOI for digital datasets was registered in 2004. Originally developed with the purpose of providing permanent access to static datasets referred to in scholarly literature, DOIs are increasingly being used for dynamic datasets, collections of datasets, and networks (e.g. time series from observational networks) (Elger *et al*., 2020). Persistent identifiers, such as DOIs, are ideal tools for providing citable and traceable references to various types of sources (e.g. data, software, samples, equipment), and importantly, it is a means towards rewarding the originators of the data. DOI-referenced and cited datasets comply with the FAIR data principles of Wilkinson *et al*. (2016).

DOI-referenced and cited datasets are (Elger, 2021a):

- **Findable** on the Internet, in repositories, databases, etc. (e.g., CDDIS and UNAVCO)
- **Accessible** via the DOI link (e.g., https://doi.org/10.prefix/suffix)

- **Interoperable** with applications or workflows
- **Reusable** – resolve to DOI landing page, containing standardized metadata and documentation.

Fig. 1 FAIR data principles (Wilkinson *et al.*, 2016).

DOI-referenced and cited datasets are also:

- Machine readable,
- Exchangeable – standardized metadata enhance data discovery via Online Public Access Catalogues (OPACs),
- Trackable in scholarly literature, and
- Used for acknowledging and rewarding institutions and researchers.

3 Initial Questions Regarding DOIs for IVS

- What type of data and products do the IVS have (i.e., data classification)?
- Who is already using DOIs and for which data, products, etc.?
- How is the scientific community using IVS data and products?
- Where do we find best practices for citable data and products?
- What about unstandardized metadata and the many schemes available?
- Should DOI-related metadata be included in existing standards, or should it remain separate and complementary to existing standards?
- Who will be responsible for minting DOIs for the IVS (licensing, etc.)?
- What about granularity of data and products? Should DOIs be assigned for instruments, stations, and networks of the IVS community?

The scientific community's response to the question of granularity of data and data products and whether DOIs should be assigned for instruments, stations, and networks of the IVS is summarized in Figure 2.

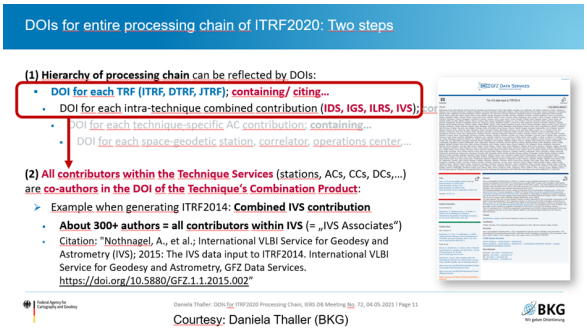

Fig. 2 Hierarchy of a processing chain reflected by DOIs (Thaller, 2021).

4 IVS Components, Data and Products

An inventory of IVS components, data, and products was conducted. This step is crucial for the DOI attribution process. It was ascertained that the IVS consists of various components as depicted in Figure 3 and that both static and dynamic data are generated by these IVS components. Products are derived from the data.

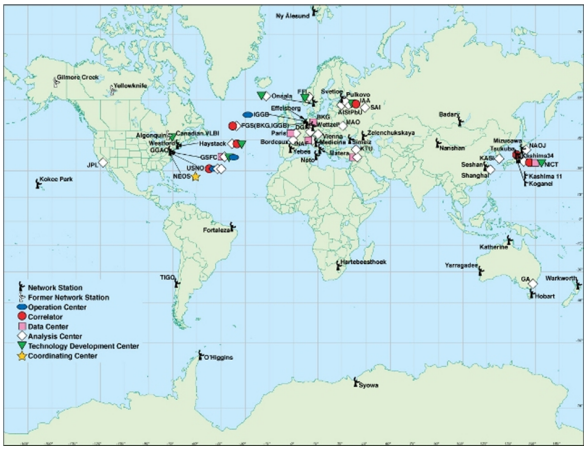

Fig. 3 IVS components (IVS, 2022).

4.1 IVS Data (Static and Dynamic)

- DB: Database files, binary fields in vgosDB format (e.g. ivsdata/vgosdb_bonn/)
- NGS: NGS card files, ASCII files in NGS card format (ivsdata/ngs/)
- SWIN: SWIN files, binary files of fringe visibilities in DIFX SWIN format (ivsdata/swin)
- AUX: Auxiliary files, ASCII files including schedules, notes and session log files (ivsdata/aux/)

4.2 IVS Products

- EOP-I: Intensive EOP series (ivsproducts/eopi/)
- EOP-S: Session EOP series (ivsproducts/eops/)
- TRF: Terrestrial Reference Frame (ivsproducts/trf/)
- CRF: Celestial Reference Frame (ivsproducts/crf/)
- DSNI: Daily UT1 solutions (ivsproducts/int_sinex/)
- DSNX: Daily EOP + station coordinates solutions (ivsproducts/daily_sinex/)
- TROP: Tropospheric parameters (ivsproducts/trop/)
- BASELINE: Time series of baseline lengths (available only on BKG/DGFI Combination Center Web site)
- VTRF: Station positions and velocities results from accumulated station coordinates, from combined normal equations of any 24-hour session (BKG/DGFI Combination Center)

5 Use Cases

To determine what other communities are doing regarding DOI minting, use cases were reviewed.

5.1 The Global Geodetic Observing System (GGOS) Working Group on DOIs for Geodetic Datasets

This Working Group (WG) was established in 2019 by the International Association of Geodesy's (IAG) GGOS and is chaired by Dr. Kirsten Elger (GFZ, Potsdam). Objectives of the GGOS WG are to (Elger, 2021a):

- address challenges and identify opportunities for improved coordination for the use of DOIs within the geodetic community,
- explore DOI minting and citation practices from other communities,
- establish best practices,
- advocate for consistent implementation of DOIs across the IAG and the geodetic community,
- explore geodetic metadata – standards, e.g. GeodesyML, as well as other DOI related discovery metadata, and
- explore the inclusion of Persistent Identifiers (PIDs), e.g., Open Researcher and Contributor Identifiers (ORCIDs) and Research Organization Registry (ROR).

5.2 NASA Earth Science Data and Information System (ESDIS)

This project started investigating the assignment of DOIs to data products in 2010 (Wanchoo, James, and Ramapriyan, 2017). Objectives of ESDIS are as follows:

- developing processes, guidelines, and models for creating and assigning DOIs
- capturing of product metadata
- establishing an automated system for assignment of DOIs

5.3 International Federation of Digital Seismograph Networks (FDSN)

This organization developed the concept of network DOIs used for citation purposes. The FDSN is responsible for (Clark, Evans and Strollo, 2014):

- providing DOI services, which include DOI mapping, minting and management services,
- promoting the use of DOIs by all networks using FDSN-assigned network codes,
- recommending metadata fields based on the DataCite model,

- providing citation examples for 'self-minted' DOIs and 'FDSN-minted' DOIs, and
- hosting the FDSN-minted DOI landing page (http://www.fdsn.org/networks/detail/);
  Other network operators who mint DOIs are responsible for their own landing pages (e.g., http://geofon.gfzpotsdam.de/doi/network/GE).

6 Metadata

An identifier, such as a DOI, is of no value without some related metadata describing what is being identified. DataCite is a leading organization devoted to promoting better access to research data. DataCite's metadata scheme is one of a number of metadata schemes available to the public. The FDSN community strongly encourages the use of DataCite's metadata model.

Metadata concepts to be considered by the IVS:

- DataCite DOI metadata scheme, ISO 19115 metadata standard
- Mandatory and discovery properties for metadata (see Figure 4)
- GeodesyML and PIDs (e.g., ROR, ORCID, FundRef, etc.)
- Controlled vocabularies
- Suffix-naming conventions and namespacing for next-consecutive-integer DOI naming that is meaningful and machine-readable
- Machine-readable embedded metadata (e.g., using <meta> tags in landing pages)

7 Citation

To counter the stigma that geodesy researchers produce less countable scientific output, it is recommended that landing pages of networks, institutions, etc. include a *'Cite this Dataset'* which should feature a pre-generated reference (see Figure 5) that users can copy-paste (Elger, 2021a). Thus, citing the resource/data is easy and guaranteed to include the DOI (Elger, 2021b).

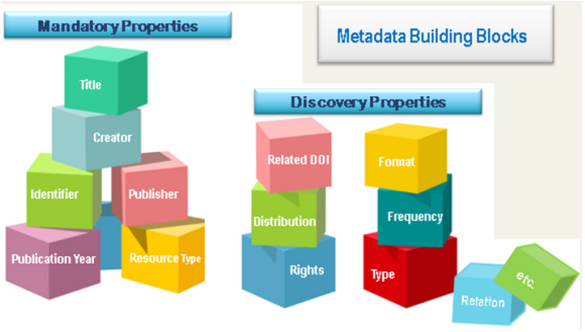

Fig. 4 Metadata properties (Bordelon, Grothkopf and Meakins, 2018).

Food for thought ...

- Citing all sources of data is good scientific practice and required by the *CC BY 4.0 Licence*.
- Citation of final products should include existing DOIs related to these products.
- DOIs of related work need to be cited (see Elger, (2021b) for INTERMAGNET Global Magnetic Observatory use case).

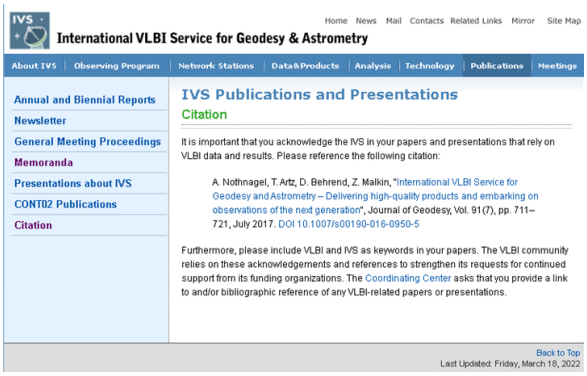

Fig. 5 Example of a pre-generated citation of a paper (IVS, 2022).

8 Conclusions

To adhere to the FAIR data principles, it is essential to attribute DOIs to IVS data and products. DOIs provide a structured and well-documented mechanism which enhances citability and scientific recognition. Cur-

rently, community-agreed methods for assigning DOIs and the application of a uniform metadata standard are lacking. Going forward, several questions need answering first. Community agreement is required to address issues such as: responsibility for minting of IVS DOIs (agency, client, etc.), metadata standards, homogenized approaches on completing metadata (with practical examples), and citation formats. Also required is the provision of DOI services (e.g. seismology community use case) allowing data users to enquire if an IVS dataset is associated with a DOI (harvest DOIs and associated metadata) and whether there are any restrictions/embargoes and licences. Knowledge gained from the GGOS WG on DOIs for Geodetic Datasets can be put to good use in our endeavor to attribute DOIs to IVS data and products.

Acknowledgments

We thank Dirk Behrend and Axel Nothnagel for sharing their insights and knowledge with us. We wish to also thank Kirsten Elger and the GGOS WG on DOIs for Geodetic Datasets for the work they are doing and from whom much of the information presented in this poster originates.

References

1. Bordelon, C., Grothkopf, U. and Meakins, S. 2018. First light for DOIs at ESO. Library and Information Services in Astronomy VIII. European Physical Journal Web of Conferences. EDP Science.
2. Clark, A., Evans, P.L. and Strollo, A. 2014. FDSN recommendations for seismic network DOIs and related FDSN services. https://www.fdsn.org/pdf/V1.0-21Jul2014-DOIFDSN.pdf
3. Elger, K. 2021a. News from the GGOS DOI Working Group. GGOS Days 2021. October 11–13, 2021. https://ggos.org/event/ggos-days-2021/
4. Elger, K. 2021b. DOI minting and citation strategies for complex/hierarchical geodetic products: example GNSS products. GGOS WG Meeting presentation.
5. Elger, K., Coetzer, G., Botha, R. and GGOS DOI Working Group. 2020. Why do geodetic data need DOIs? First ideas of the GGOS DOI Working Group. EGU General Assembly 2020. https://doi.org/10.5194/egusphere-egu2020-17861
6. IVS. 2022. IVS Organizations: Components. https://ivscc.gsfc.nasa.gov/about/org/components/index.html
7. Thaller, D. 2021. DOIs for ITRF2020 processing chain. IERS DB Meeting, 72. 4 May 2021.
8. Wanchoo, L., James, N. and Ramapriyan, H.K. 2017. NASA EOSDIS Data Identifiers: approach and system. *Data Science Journal*, 16:15, 1–11.
9. Wilkinson, M.D. et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*, 3:160018. https://doi.org/10.1038/sdata.2016.18