

Identification and citation in IAGOS

Context of identification and citation in IAGOS

Summary of IAGOS requirements for identification and citation

Detailed requirements

1. Identification

i. What granularity do your RI's data products have:

- **Content-wise (all parameters together, or separated e.g. by measurement category)?** all parameters together, download one or several parameters
- **Temporally (yearly, monthly, daily, or other)?** flight (at least daily)
- **Spatially (by measurement station, region, country or all together)?** global (flight trajectory)

ii. How are the data products of your RI stored - as separate "static" files, in a database system, or a combination? database system

iii. How does your RI treat the "versioning" of data - are older datasets simply replaced by updates, or are several versions kept accessible in parallel? How do you identify different version of the same dataset ? since IGAS we keep all versions for archive but only last available to users

iv. Is it important to your data users that

- **Every digital data object is tagged with a unique & persistent digital identifier (PID)?** yes
- **The metadata for data files contains checksum information for the objects?** no
- **Metadata (including any documentation about the data object contents) is given its own persistent identifier?** file identifier (ISO 19115)
- **Metadata and data objects can be linked persistently by means of PIDs?** on going

v. Is your RI currently using, or planning to use, a standardized system based on persistent digital identifiers (PIDs) for:

- **"Raw" sensor data?** no
- **Physical samples?** no
- **Data undergoing processing (QA/QC etc.)?** no
- **Finalized "publishable" data?** yes

vi. Please indicate the kind of identifier system that are you using - e.g. Handle-based (EPIC or DOI), UUIDs or your own RI-specific system? DOI

vii. If you are using Handle-based PIDs, are these handles pointing to "landing pages"? Are these pages maintained by your RI or an external organization (like the data centre used for archiving)? by French Atmospheric Data Center AERIS

viii. Are costs associated with PID allocation and maintenance (of landing pages etc.) specified in your RI's operational cost budget? In kind

2. Citation

i. How does your "designated scientific community" (typical data users) primarily use your data products? As input for modelling, or for comparisons?

comparison with satellite or model for validation, atmospheric process studies and trends analysis

ii. Do your primary user community traditionally refer to datasets they use in publications:

- **By providing information about producer, year, report number if available, title or short description in the running text (e.g. under Materials and Methods)?** yes
- **By adding information about producer, year, report number if available, title or short description in the References section?**
- **By DOIs, if available, in the References section?** yes
- **By using other information?**
- **By providing the data as supplementary information, either complete or via a link**

iii. Is it important to your data users to be able to refer to specific subsets of the data sets in their citation? Examples:

- **Date and time intervals**
- **Geographic selection**
- **Specific parameters or observables**

yes to all

iv. Is it important to be able to refer to many separate datasets in a collective way, e.g. having a collection of "all data" from your RI represented by one single DOI? yes

v. What strategy does your RI have for collecting information about the usage of your data products?

- **Downloads/access requests** yes
- **Visualization at your own data portal** yes

- **Visualization at other data portals** no
- **References in scientific literature** yes
- **References in non-scientific literature** no
- **Scientific “impact”** yes (number of citation)

vi. Who receives credit when a dataset from your RI is cited?

- The RI itself
- The RI's institutional partners (all or in part, depending on the dataset contents) yes
- Experts in the RI's organization (named individuals) yes
- “Principal investigators” in charge of measurements or data processing (named individuals) yes
- Staff (scientists, research engineers etc.) performing the measurements or data processing (named individuals) yes

What steps in tooling, automation and presentation do you consider necessary to improve take up of identification and citation facilities and to reduce the effort required for supporting those activities?

ask to editors to check RI citation requirements for new papers

Formalities (who & when)

Go-between	Yin Chen
RI representative	Damien Boulanger < damien.boulanger@obs-mip.fr >
Period of requirements collection	Nov-Dec 2015
Status	Completed