

Identification and citation in SIOS

Context of identification and citation in SIOS

Summary of SIOS requirements for identification and citation

Detailed requirements

Identification

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

- **Content-wise (all parameters together, or separated e.g. by measurement category)?** sensor ,, measurements category , time
- **Temporally (yearly, monthly, daily, or other)?** yearly, montly, daily, possibly hourly
- **Spatially (by measurement station, region, country or all together)?** measurement area/region, measurement station,

2. How are the data products of your RI stored - as separate "static" files, in a database system, or a combination? they can be stored separately as well as in a database

3. How does your RI treat the "versioning" of data - are older datasets simply replaced by updates, or are several versions kept accessible in parallel? [can't answer]

4. 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? yes
- Metadata (including any documentation about the data object contents) is given its own persistent identifier? yes
- Metadata and data objects can be linked persistently by means of PIDs? yes

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

- "Raw" sensor data? this will be decided/discussed during the implementation phase
- Physical samples? as above
- Data undergoing processing (QA/QC etc.)? as above
- Finalized "publishable" data? as above

6. 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? can't answer

7. 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)? can' t answer

8. Are costs associated with PID allocation and maintenance (of landing pages etc.) specified in your RI's operational cost budget? can' t answer

Citation

1. How does your "designated scientific community" (typical data users) primarily use your data products? As input for modelling, or for comparisons? testing theory, comparison model result, process analysis

2. 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? yes
- By DOIs, if available, in the References section? yes
- By using other information? can't answer

3. 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 yes
- Geographic selection yes

- Specific parameters or observables yes

4. 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? can't answer

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

- Downloads/access
- Visualization at your own data portal
- Visualization at other data portals
- References in scientific literature
- References in non-scientific literature
- Scientific “impact”

they could all used

6. 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)
- Experts in the RI's organization (named individuals)
- “Principal investigators” in charge of measurements or data processing (named individuals)
- Staff (scientists, research engineers etc.) performing the measurements or data processing (named individuals)

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?

All of the point listed can have credits, a standard way should be defined when metadata are provided.

Formalities (who & when)

Go-between	(yinchen)
RI representative	Angelo Viola
Period of requirements collection	Dec 2015- Jan 2016
Status	Complete