Data Citation Service: Workflow

Data Citations are provided on Model/MIP and experiment granularities.

Workflow for Modeling Centers:
1. Contact Person for citation required
2. Provide at least basic data citations for model data: author lists + titles via GUI
3. First data delivered to ESGF → DataCite DOI registered
4. Update model/MIP data citations and add citations for experiment data via API → automated registration of updated metadata at DataCite
5. At time of long-term archival (AR6 WG I literature-cutoff date: 2020-10-15), citation information will be fixed.
Data Citation Service: Concept
Data Citations Services uses ESGF for data access.

CMIP6 data citation concept:

- WDCC/DKRZ hosts the repository and the landing page
- ESGF provides the data access
- CMIP6 data citations use DataCite DOIs issued by WDCC on behalf of the ESGF; publisher is the ESGF. (reuse of existing DOIs registered by data centers is possible)
- Stable citation reference with version information to identify data status and ESGF support filtering for version ranges:

  Authors (YYYY): Title. **Version YYYYMMDD**. Earth System Grid Federation. DOI.

Matthes, Katja; Funke, Bernd; Kruschke, Tim; Wahl, Sebastian (2017). input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2. **Version 20170103**. Earth System Grid Federation. [https://doi.org/10.22033/ESGF/input4MIPs.1122](https://doi.org/10.22033/ESGF/input4MIPs.1122)
CMIP6 data citation components:

- Oracle Database as repository
- Oracle APEX GUI to maintain citation information
- Metadata disseminated as DataCite 4.0 in XML and JSON formats via API using content negotiation
- XMLs accessible via OAI server

CMIP6 data citation integration:

**Within CMIP6:**
Citation information will be accessible in the ESGF portals and Citation URLs (API JSON) included in ESGF Solr index

**Outside CMIP6:**
The citation information registered with DataCite includes ORCIDs and CrossRef DOIs for literature to enable interlinking. It is planned to use the currently developed Scholix (http://www.scholix.org) service to request data references in literature.
Data Citation Service: Information and Status (1)

Data Citations are provided on Model/MIP and experiment granularities.

Information and Status – technical perspective:

- Specifications at: http://bit.ly/1XsVOoz

**Oracle APEX GUI** is operational:
- No. of citation managers (2017-11-10): CMIP6:21; input4MIPs:13
- API for provision of citation information is in test phase
  → release planned for early 2018
- **Landing page** with dynamic content is operational,
  e.g. https://doi.org/10.22033/ESGF/input4MIPs.1122
- **ESGF CoG** implementation is done

Other Background Functionalities:
- Insert information from CV is a manual process
- Registration of DataCite metadata/DOI is a manual process
  → will be automated after first CMIP6 data is published
- DataCite metadata update and integration in set on OAI server are automated
- Integration in furtherInfo URL landing page is under discussion with ES-DOC
- Metadata curation needs to be improved and partly automated
Data Citation Service: Information and Status (2)

Data Citations are provided on Model/MIP and experiment granularities.

Information and Status - user perspective:

- General information: [http://cmip6cite.wdc-climate.de](http://cmip6cite.wdc-climate.de)
- Citation service used for input4MIPs, e.g. [https://doi.org/10.22033/ESGF/input4MIPs.1122](https://doi.org/10.22033/ESGF/input4MIPs.1122)
  → documentation to be improved; web demo planned for 2018
- API Tool to provide information on fine granularity in test phase (release planned for early 2018)
- Data Citation accessible in ESGF CoG portal (after republication of input4MIPs data)
Long-Term Archival
Transfer CMIP6 snapshot into IPCC DDC

Information and Status:

• Snapshot date is 2020-10-15
• IPCC DDC AR6 data will receive DataCite DOIs
  with documented relations to the DOIs for evolving CMIP6 data ("early citation")

✅ Data replication is clear → ESGF-RVWT

⚠️ Integration of ancillary metadata as result of ESGF-QCWT is planned
  via ESGF-registered URLs and information on content provided,
  see https://acme-climate.atlassian.net/wiki/spaces/ESGF/pages/53018883
  → currently only provided by citation service
  → integration of es-doc and errata information is currently unclear

• Linking to external ancillary metadata sources out of the IPCC DDC requires
  persistent IDs or persistent URLs for the ancillary metadata object