



EGU2015-5570

10 years of data publication with Digital Object Identifiers (DOI) at the World Data Center for Climate (WDCC)

M. Stockhause, M. Lautenschlager, H. Höck, and F. Toussaint
German Climate Computing Center (DKRZ; stockhause@dkrz.de)



Concept Development: -2002

Basic Ideas:

The idea to apply DOIs on data was first investigated in a study at GFZ in 1998 (Mundt, 1998). The following years, a concept was developed by DKRZ and GFZ at the 17th Int. CODATA Conference in 2000 (Lautenschlager & Wächter, 2000).

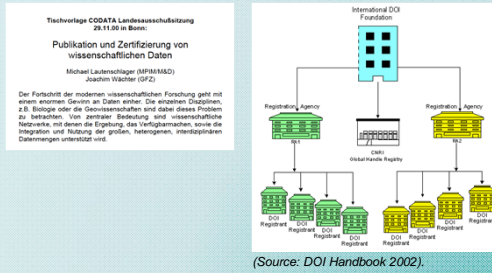
Concept and Implementation Plan:

End of 2000 a CODATA Working Group consisting of librarians and scientists was formed. DFG funded their work for two years. Their final report was published 2002, defining the concept of scientific data publication and suggesting an implementation plan (CODATA AG, 2002).

Guiding Principle for Scientific Data Publication:

Analogy to scientific literature as close as possible, which means scientific data are irrevocable after publication and have a suitable granularity.

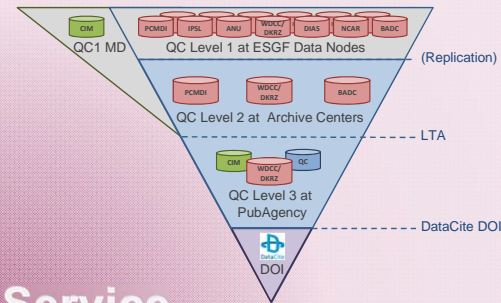
Concept at 17th Int. CODATA Conference



Concept

2009-2014: Becoming a Service

Quality Assurance Process for CMIP5/IPCC-AR5



Service

Operability:

For the operability of WDCC's DOI data publication process - the process was automated and became a service, - a GUI for metadata generation was created, - data citation was integrated in WDCC's data portal and in metadata provided for other community portals.

Flexibility:

WDCC integrated the DOI publication process into different long-term archival (LTA) procedures, e.g. it became part of the overall quality assurance process for CMIP5 / IPCC-AR5 data. Possible ways to a data DOI at WDCC / DKRZ are: - LTA of project data from DKRZ's HPC environment, - LTA of external project data, - LTA of project data disseminated in the ESGF.

DataCite as international organization, formed in 2009, resumes the data publication engagement (<http://datacite.org>).

Implementation: 2003-2008 Realisation

Concept:

Quality controlled scientific data are classified as irrevocable and registered as DOI (URN) with citation reference in a CNRI Handle Server and in library catalogs in order ...

- To allow for transparent data access
- To foster verification of scientific results
- To allow for data citation in scientific literature
- To give credit to data authors

Implementation:

TIB and 3 data archives, coordinated by WDCC, implemented the STD-DOI data publication system with the components:

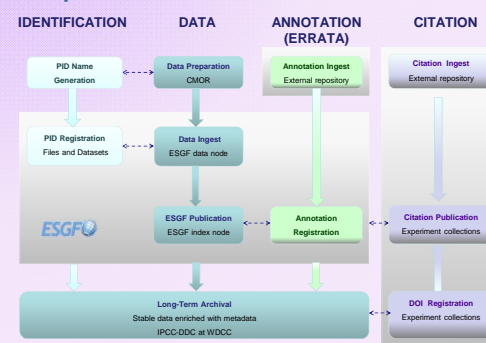
- The STD-DOI metadata scheme (cross-disciplinary)
- STD-DOI registration and catalog services at TIB

The definition of data granularity across disciplines suitable for citations in scientific literature turned out to be difficult.

STD-DOI Landing Page and Catalog Entry

Future 2015- : Service Integration

Concept for Data Citations for CMIP6/IPCC-AR6



Integration in Research Environment:

As integral part of the future research environment, data DOIs will be connected via Persistent Identifiers (PID): - to PIDs on files and datasets incl. replica, - to PIDs on non-data objects like scientific publications, annotations (errata, user annotations), data creation method (model, simulation, experiment) - to researchers' person IDs, e.g. ORCID

Data Citation Concepts for CMIP6:

WDCC takes the lead in the data citation concept for CMIP6. The focus for improvements lies in: - Data citation information for dynamic project data to provide early data citations, and - Improved visibility and accessibility of citation information for the data objects as well as paper publications based on CMIP6 data.

References and additional information:

References:

- J. Brase, M. Lautenschlager, I. Sens (2015): *The Tenth Anniversary of Assigning DOI Names to Scientific Data and a Five Year History of DataCite*. D-Lib Magazine. Volume 21. Number 1/2. doi:10.1045/January2015-brase.
- M. Stockhause, H. Höck, F. Toussaint, M. Lautenschlager (2012): *Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data*. Geosci. Model Dev., 5, 1023-1032, 2012, doi:10.5194/gmd-5-1023-2012.

Materials:

- M. Mundt (1998): *Der DOI (digital object identifier), ein verlagsorientiertes Indexierungswerkzeug auch anwendbar auf Datensätze?* GFZ.
- M. Lautenschlager, J. Wächter (2000): *Publikation und Zertifizierung von wissenschaftlichen Daten*. 17th CODATA Conference.
- CODATA AG "Possibility of citing scientific primary data" (2002): *Conception of citing scientific primary data, Final Report*.

Information:

- WDCC Data Publication Process: <http://www.dtkz.de/daten-en/data-services/Datapublication>
- CMIP5 Data Publication Process: <http://cmip5qc.wdc-climate.de>
- IPCC DDC AR5 Data Citations: http://www.ipcc-data.org/sim/gcm_monthly/AR5
- STD-DOI / KomFor: <http://www.std-doi.de>
- DataCite: <http://datacite.org>