



Citing Dynamic Data - Research Data Alliance working group recommendations

Ari Asmi (1,5), Andreas Rauber (2), Stefan Proßl (3), and Dieter van Uytvanck (4)

(1) University of Helsinki, Physics, Helsinki, Finland (ari.asmi@helsinki.fi), (2) Vienna University of Technology, Austria (rauber@ifs.tuwien.ac.at), (3) SBA Research, Vienna, Austria (sproell@sba-research.org), (4) CLARIN ERIC, Utrecht, Netherlands (dieter@clarin.eu), (5) ICOS ERIC, Helsinki, Finland

Geosciences research data sets are typically dynamic: changing over time as new records are added, errors are corrected and obsolete records are deleted from the data sets. Researchers often use only parts of the data sets or data stream, creating specific subsets tailored to their experiments. In order to keep such experiments reproducible and to share and cite the particular data used in a study, researchers need means of identifying the exact version of a subset as it was used during a specific execution of a workflow, even if the data source is continuously evolving.

Some geosciences data services have tried to approach this problem by creating static versions of their data sets, and some have simply ignored this issue. The RDA Working Group on Dynamic Data Citation (WGDC) has instead approached the issue with a set of recommendations based upon versioned data, timestamping and a query based subsetting mechanism.

The 14 RDA WGDC recommendations on how to adapt a data source for providing identifiable subsets for the long term are:

Preparing the Data and the Query Store

R1 - Data Versioning

R2 - Timestamping

R3 - Query Store Facilities

Persistently Identifying Specific Data Sets

R4 - Query Uniqueness

R5 - Stable Sorting

R6 - Result Set Verification

R7 - Query Timestamping

R8 - Query PID

R9 - Store the Query

R10 - Automated Citation Texts

Resolving PIDs and Retrieving the Data –

R11 - Landing Page

R12 - Machine Actionability

Upon modifications to the Data Infrastructure

R13 - Technology Migration

R14 - Migration Verification

We present a detailed discussion of the recommendations, the rationale behind them, and give examples of how to implement them.