CCCA Dynamic Data Citation, an Extension on Earth Observation Data

The Climate Change Centre Austria, as a research data infrastructure facility for the Climatological Domain, targets a very limited user community. With the support of an RDA Europe 4.0 grant, the project team, led by Chris Schubert, has implemented the Scalable Dynamic-data Citation Methodology in order to extend the CCCA's services to a further research domain, namely the Earth Observation Sciences. This project follows on from earlier work carried out during the RDA Europe 3.0 project and documented in a previous Adoption Story.

The challenge

The Climate Change Centre Austria (CCCA) has fully implemented the Recommendation of the RDA Working Group on Data Citation. The search for an opportunity to apply the Dynamic Data Citation Service to different domains, in this case combining satellite data and climate models, has brought us to this challenge. The technical challenges that we needed to solve involved the differing technological stacks and concepts used by CCCA and the Earth Observation Data Centre (EODC), as well as the challenge of combining climate data and earth observation data. The main design challenge was to supply data citation for the providers by making use of existing architectures and concepts (e.g. Austrian Data Cube, openEO) and to construct a generic solution with the potential for usage growth.

The RDA outputs adopted

The Recommendations of the RDA Working Group on Data Citation: Data Citation of Evolving Data (Rauber et al. 2015) were adopted by the CCCA. The same output was used for gaining data citation for EODC and CCCA, since these Recommendations have already a history of approval and successful adoptions in both organisations.

Benefits of adoption and impact

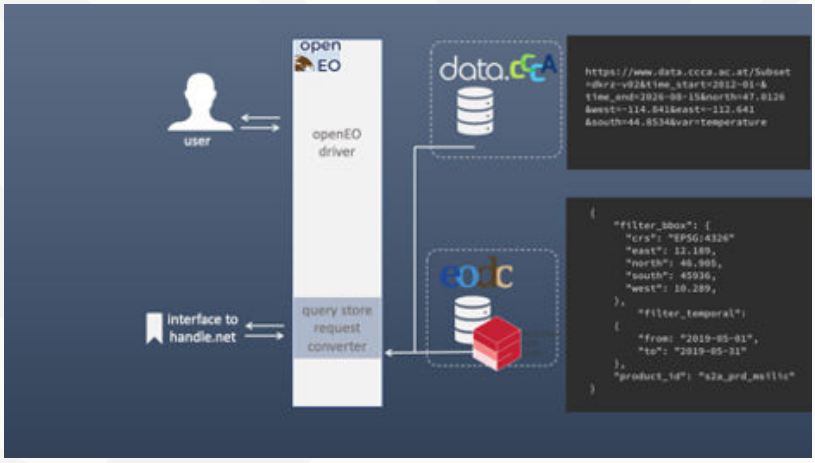
As CCCA data centre, initiated for a small community like the climate scientists in Austria, the end users did not initially recognise the value of citing data. Two years after the release of this dynamic data citation service, the need changed into acknowledgement and acceptance. Nevertheless, there is still some work to be done to ensure that data providers internalise the dynamics of research data and their publication.

RDA and the Data Citation WG promoted the technical approach to implementation and gave us the scientific backbone. But in 2019-2020, the user needs changed to a need for precise information of data fragments, as well as evolving data at specific times regarding trustful and reproducible data and methods. This also signalled the need for a data repository that is able to lower the barrier to data publishing by enabling a clear method to keep the history of the data.

Adopting the Data Citation WG Recommendations in a combined effort by CCCA and EODC has enabled the citation of subsets between scientific areas. It makes it easier for researchers to cite used data, since they do not have to change the way they work, and they can get persistent identifiers for data subsets. The combined effort improves the workflow of scientists by providing a common cross-scientific method on data citation.

The adoption process

The CCCA, with the assistance of TU Vienna, started researching the in-place data citation of the climate and earth observation providers, with a focus on the Copernicus Climate Change Service. The technological overview of EODC and CCCA in the context of data retrieval was arranged and was the starting point for the design process. Further investigations were carried out into the potential of ACube and openEO for integrating the RDA Recommendations, resulting in several concepts to connect the data provider. The most promising was the application of an existing generic openEO driver as a connector for different data repositories, which could then be used to ease up the data citation process based on the 14 Recommendations of the Data Citation Working Group. CCCA, as well as EODC, can use this generic driver, which enables the request to data catalogues via a Persistent Identifier, and obtain the meta-information which is needed as a software module on top of their (combined) infrastructure.



Lessons learned

This adoption process at this time period in 2020 shows us the role of importance of running data infrastructures, their vulnerabilities, fears and threats, and was characterised by the fact that real work on existing software systems shows a lot of dependencies, e.g. security, available effort, resources, etc.

The personal lesson learned is that we now have the impression that there is no infrastructure of data providers in which data citation cannot be achieved with the help of the RDA Recommendations, and the outcome of this Adoption Story strengthens this further.

About CCCA

The CCCA Data Centre as an operational department of the Climate Change Centre Austria (CCCA) is responsible for the research data infrastructure to promote applicable data management for Austrian researchers and the Greater Alpine Region. Our main objective is to provide a central climate data hub for models, climate scenarios, related research data and information services. The CCCA Data Centre fosters data sharing principles and standardized data services for download, view and data analyzing tools. Together with the expertise of the Technical University Vienna (TU Wien), Department of Geodesy and Geoinformation on operational services with Sentinel data as well as using High Performance Computing (HPC) in a operational sense, we have the same view on the importance on the topic on citation of evolving data. A third organization was involved as an infrastructure platform, the Earth Observation Data Centre (EODC).

Climate Change Centre AUSTRIA

Find out more at:

www.rd-alliance.org/recommendations-outputs

Visit rd-alliance.org

or write us at enquiries@rd-alliance.org