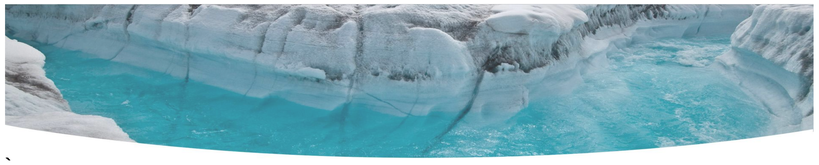
Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Enabling data-driven provenance in NetCDF, via OGC WPS operations. Climate Analysis services use case.

Andrej Mihajlovski, Alessandro Spinuso, Maarten Plieger and Wim Som de Cerff *The Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands*



`

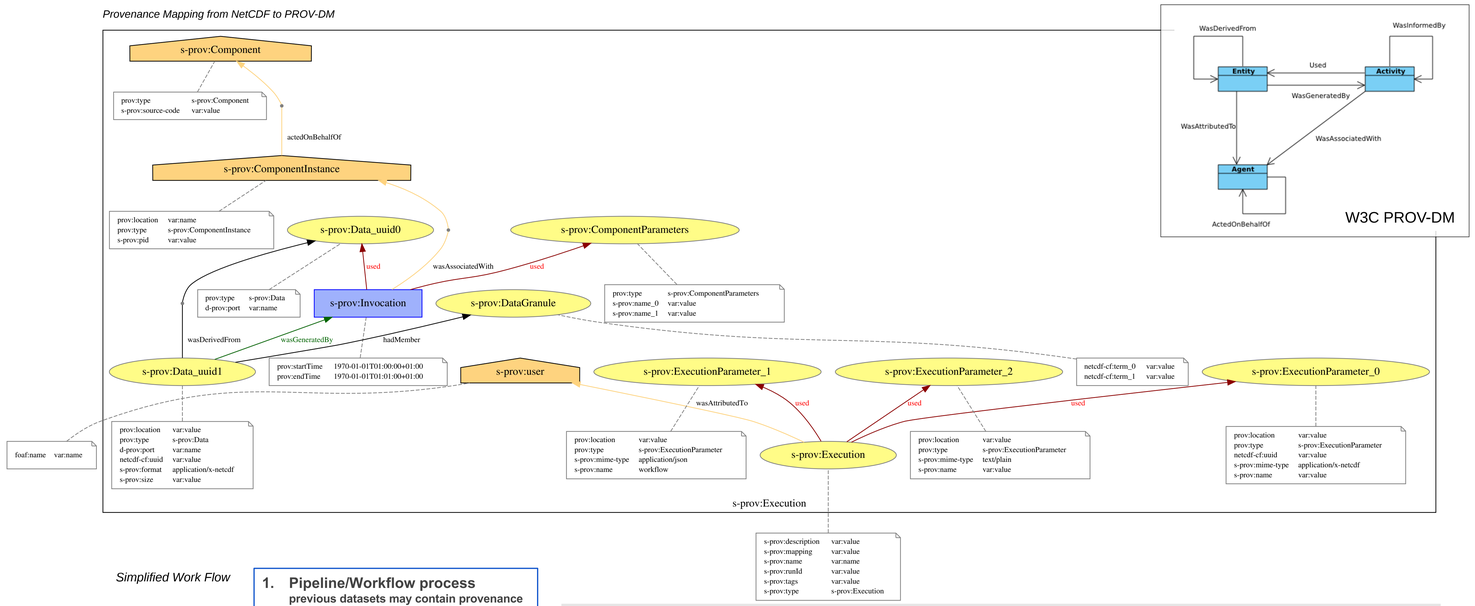
USE CASE: Copernicus Climate Information Portal (CLIPC) Processing Toolbox

Copernicus is a European system for monitoring the Earth. CLIPC is developing pre-operational Copernicus Climate Change services (C3S)

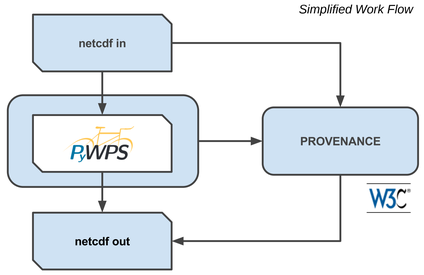
Modern Climate analysis platforms provide generic and standardized ways of accessing data and processing services. These are typically supported by a wide range of OGC services and interfaces, such as the Web Processing Service (WPS). However, the problem of instrumentally tracing the lineage of the transformations occurring on a dataset and its provenance remains an open challenge. It requires standard-driven and inter-operable solutions to facilitate understanding, sharing of self-describing data products, fostering collaboration among peers.

The CLIPC portal provided us with a real use case, where the need of an instrumented provenance management is fundamental. CLIPC provides a single point of access for scientific information on climate change. The data about the physical environment which is used to inform climate change policy and adaptation measures comes from several categories: satellite measurements, terrestrial observing systems, model projections and simulations and from re-analyses.

Provenance Mapping from NetCDF to PROV-DM



Simplified Work Flow



1. **Pipeline/Workflow process**
   previous datasets may contain provenance encoding which is integrated into current PROV.
2. **Metadata snapshot**
   ncattributes and variable attributes extracted from input NetCDF
3. **Provenance**
   a. Bundle
      Process (Activity) attributes
      User/Software (Agent)
      Links consecutive processes
   b. Lineage
      Metadata Content
      Process Work Flow Described
   c. Lineage of Input NetCDF added
4. **Output**
   provenance variable added to netCDF
   a. Bundle (json)
   b. linage (json)
   c. Prov-dm (xml)

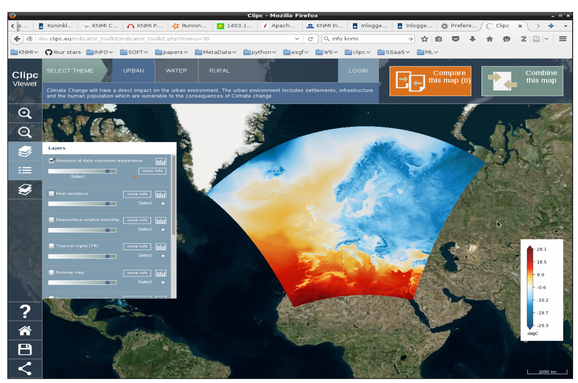
Core contributions made

1. Provenance wrapper around WPS services
2. Provenance tracing within the NetCDF format, which adopts and extends the W3C's PROV model.
3. Define a structure for a NetCDF provenance variable
4. Disseminate climate indicator data and create transformed data products
5. Standardized provenance, metadata and processing infrastructure is researched for CLIPC.

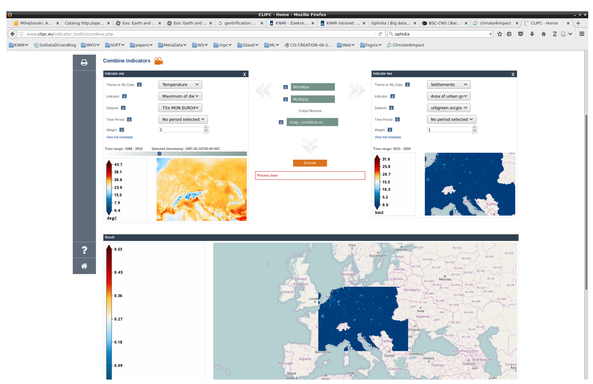
These efforts will lead towards the provision of tools for further web service processing development and optimization, opening up possibilities to scale and administer abstract users and data driven workflows.

Portal Tool Kit

The climate impact tool kit is realized with the orchestration of a number of WPS that ingest, normalize and combine NetCDF files. The WPS allowing this specific computation are hosted by the climate4impact portal, which is a more generic climate data-access and processing service. In this context, guaranteeing validation and reproducibility of results, is a clearly stated requirement to improve the quality of the results obtained by the combined analysis.



CLIPC Portal: Climate Toolbox



CLIPC Portal: Combine Process Interface

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418.

KNMI CLIPC TEAM

clipc@knmi.nl
Royal Netherlands Meteorologic Institute
www.climate4impact.eu

www.clipc.eu