

Observations closer to Model Data (OcMOD) -

Frank Kaspar¹, Ivonne Anders², Christopher Kadow², Joaquim Pinto³, Kevin Sieck⁴

¹ Deutscher Wetterdienst, DWD, Offenbach

² Deutsches Klimarechenzentrum, DKRZ, Hamburg

³ Karlsruher Institut für Technologie, KIT, Karlsruhe

⁴ Climate Service Center Germany, GERICS, Hamburg

2020/07/29

Abstract

Working with climate model output nearly always include a validation of this data compared to reference data from observations, to ensure, that the model data chosen is suitable for the individual research question or application. Model data and commonly used observational data have to be obtained from different sources: model data from e.g. the DKRZ and observational data from e.g. servers from public authorities. All data have to be prepared to be in the same formats and standards before it can be used for further analysis. A broad range of users (e.g. climate modelers, impact modelers, climate scientists, providers for climate services and education) are dealing with same the effort and troubles with the same sets of data. The aim of the pilot is to bring observational data close to the model output, to easy access data from public authorities and increase the number of users of various disciplines, and to provide this data in standardized formats for easy usage. As stakeholders, we choose two big national climate change and climate impact projects, both need observations and climate model data for evaluation and climate change analysis. The pilot will integrate the DWD reanalysis dataset COSMO-REA6 (Kaspar et al. 2020) to the infrastructure of ESGF (Earth System Grid Federation - <https://esgf.llnl.gov/>) as one example. After the end of the pilot, the developed workflow will work out as a base to integrate further datasets from DWD or other public authorities.

I. Introduction

Climate model output as input for a high variety of research, climate impact analysis and adaptation strategies are therefore of high value for the society's future. Global availability of large volume climate data is achieved via the ESGF, an infrastructure of internationally distributed data nodes enabling the management, dissemination and analysis of large-volume scientific, mainly simulation/model, data. As the output data from state-of-the-art climate models is becoming too large to download and make local analysis impossible for users, server-side solutions are developed and are a big challenge for data providers.

The mentioned climate model data sets like e.g. from CORDEX (<https://cordex.org/>) or CMIP (<https://www.wcrp-climate.org/wgcm-cmip>) are available via ESGF and thereby reach a huge number of users from different disciplines. Most of the researchers and other users use model data together with observations for their specific application. The areas of application are diverse and range from changes in the groundwater level, variations in the trajectory of migratory birds, changes in the high atmosphere or permafrost to the assessment of suitable adaptation measures in e.g. coastal areas or cities. An example of very diverse research is the cluster of excellence

CLICCS of the Hamburg University, which combines almost 100 researchers, in majority dealing with modelling the impact of climate change on certain parts of the earth system.

The reference data, e.g. observations of national weather services, which is usually needed to test quality and applicability of the model data for specific analysis, have to be always obtained from other sources.

The two BMBF-funded projects ClimXtreme^{1,2} and RegIKlim³ are both using climate model output provided by ESGF and at the same time the high-resolution regional reanalyses COSMO-REA6 and REA2 to analyse climate change and impacts for Germany. ClimXtreme is focusing on the investigation of climate extremes in the past and future; within RegIKlim high resolution climate model simulations will be carried out in close collaboration with certain focus regions, to implement adaptation strategies.

Both projects use the evaluation framework FREVA (www.xces.dkrz.de) adapted for individual project needs, to evaluate and analyse climate model results and observational data. This framework is also successfully used in education e.g. at Freie Universität Berlin (FUB). All model data the same as observations have to be available in CMOR- (Climate Model Output Rewriter) standard, which is used in big climate Coupled Model Intercomparison Projects (CMIP5/6). Via the DKRZ infrastructure FREVA comes with direct access to the data provided by the ESGF node. For the two mentioned projects, COSMO-REA6 is an essential dataset to be standardized and available within the FREVA framework. At the moment all steps are prepared manually and with great effort of DWD. COSMO-REA6 has been intensively used in other applications too, e.g. related to the energy sector (Kaspar et al., 2020), but it **has not yet been integrated into research infrastructures** of the Earth System Research community.

Bringing observations collected by public authorities close to the climate model data, simplifies the access to all data which is required for individual analysis, and therefore also increases the number of users of the observations. There is **no existing common framework** to incorporate governmental data products into global research data distribution infrastructures.

With successful investigation, a framework is available to easier integrate other external observational dataset provided by the public authorities.

1 <https://www.fona.de/de/massnahmen/foerdermassnahmen/climxtreme.php>

2 <https://www.fona.de/de/massnahmen/foerdermassnahmen/regionale-informationen-zum-klimahandeln.php>

II. Pilot description

Aim of this pilot is to integrate a first data set provided by the German Weather Service, as a public authority, into the ESGF structure, and make it accessible for a broad range of researchers and other users from various disciplines. This pilot will investigate the whole framework and will facilitate the integration of further datasets from DWD but works also as an example to integrate data from other public authorities.

After collecting the basic information of the COSMO-REA6 data like sub datasets, file sizes, file structure, file formats, in close collaboration with the stakeholders from the two projects ClimXtreme and RegIKlim a list of first most relevant meteorological parameters will be compiled, processed and temporarily copied to DKRZ. An organizational layout will be designed for the new dataset to fit the regulations of ESGF, the Data Reference Syntax (DRS). To be able to integrate a new dataset into the ESGF as a user friendly dataset the very important standardization of the

data is needed. This is an intense iteration process, starting with designing and testing the fulfillment of the demanded standards including a quality control. To avoid delay in the pilot this process will be built on existing standards, e.g. CORDEX. For cost-efficient use of storage, the file size needs to be optimized. In the end of this process, the standardization is applied to the selected subset of COSMO-REA6 under supervision, and final quality checked. Before integration into the ESGF catalogue for the public access via the web interface, usage licenses need to be formulated. In close collaboration with DWD the legal options for long term archiving of this standardized subset of COSMO-REA6 at DKRZ will be proved. It allows a publication connected to a persistent identifier DOI. A main activity in the pilot is further to develop the described workflow (cf. V Workflow & Requested Funding) to work as a base for a sustainable concept for the broad application to integrate other observational datasets from other public authorities in the ESGF infrastructure. This concept will be published, as reference for other data providers. As the technological backbone, the ESGF node is located at DKRZ. DKRZ is one of the core partners of the ESGF and one of the leading partners in the IS-ENES network that coordinates the European contributions to the ESGF (and later EOSC). FTP server at DWD ensure data transfer to the infrastructural capacities for storing and processing Earth System Science data atDKRZ. With respect to the standards and to ensure interoperability, the existing COSMO-REA6 output will be converted to standardized netCDF and follow established CORDEX-conventions (data and metadata) and the organization constraints imposed by the ESGF.

The challenge is not only integrating the observational data into ESGF but also to create a common workflow together with the public authority DWD. Furthermore, the previously scattered ESS data (model, observations, evaluation) will be centrally available via one platform and enlarge the opportunities for the development of sustainable concepts for further applications and products. Challenges that go beyond the technical part (e.g. licenses) will be explored in order to be able to address them in future projects with public authorities.

III. Relevance for the NFDI4Earth

The stakeholder in this pilot are leading members of the two national funded projects ClimeXtreme and RegiKlim focusing on climate change assessment, research and adaptation climate strategies. But the pilot aims to support a wide range of other users, like e.g. impact modellers, climate researchers and education in providing relevant data via ESGF and making data from public authorities available in a high standardized format with easy access.

Regarding the **FAIR**-principles: **F** – In consultation with the DWD a persistent identifier (PID) to the observational data published in the ESGF will be assigned, like e.g. a DOI, .; **A** – ESGF is an established global web interface for data search, access and download.; **I** - The observational data is converted to domain-specific and widely-used data and metadata standards (based on the standards of CORDEX).; **R** – the during the pilot compiled DWD-DKRZ cooperation agreement will include appropriate usage license, domain-specific (meta)data standards and documentation of provenance. With this pilot the data publication and dissemination are covered as parts of the data life cycle. The outcome of the pilot and the further ability to integrate more and more observational datasets into ESGF supports the efficient education of students in climate research, but increases the applications of server-side analysis, development of new tools integrating model output the same as observations.

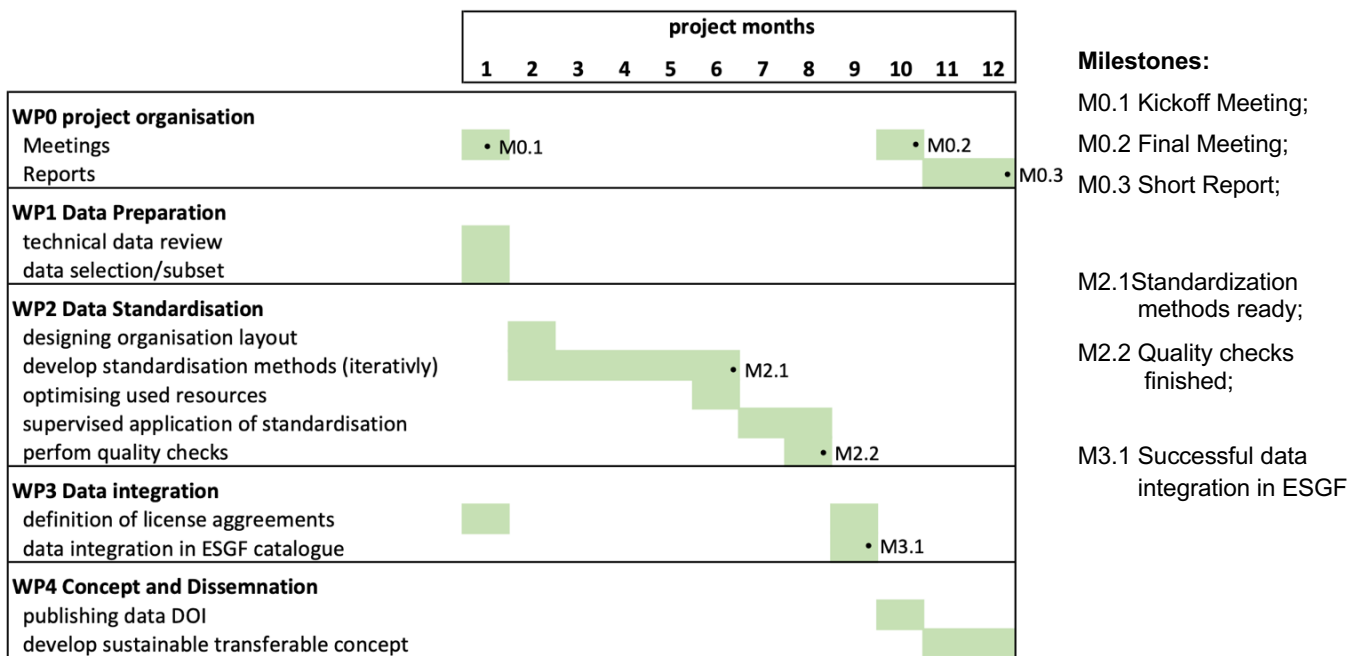
IV. Deliverables

D1: COSMO-REA6 integrated to ESGF: The dataset COSMO-REA6 is standardized and fully integrated and published in the ESGF framework and accessible via the web interface but also via direct access through DKRZ infrastructure.

D2: Basic content for a concept for the integration of observational data from public authorities: The most important basis for a sustainable concept is available, showing the whole workflow and needed steps and definition of regulations to include further observational datasets in ESGF and being usable for a broad range of applications.

V. Work Plan & Requested funding

Workplan as described in the pilot description:



Requested funding according DFG

We apply for funding of one-year full time equivalent (Postdoc, TVL13 100%) located at DKRZ in Hamburg. The work is strongly supported by DWD with short term visits to DWD Hamburg and Offenbach. DKRZ and DWD agreed that the preferred mode of contract is employment at DKRZ, because at DWD short term contracts of a few months are typically not feasible because of the additional regulations and complexity of the recruitment process. Members of ClimXtreme (C. Kadow) and RegIKlim (K. Sieck, J. Pinto) support this pilot as stakeholders and will be involved in parameter definition of the COSMO-REA6 and stay regularly informed about the progress of the data integration to be used within the FREVA framework.

Reference

Kaspar, F., et al., 2020: Regional atmospheric reanalysis activities at Deutscher Wetterdienst: review of evaluation results and application examples with a focus on renewable energy, Adv. Sci. Res., 17, 115–128, <https://doi.org/10.5194/asr-17-115-2020>, 2020.