Project: **1123**

Project title: **AtMoDat**

Principal investigator: **Hannes Thiemann**

Report period: **2020-11-01 to 2021-08-31**

We here report on work conducted within the reporting period in the framework of the AtMoDat project (www.atmodat.de), a BMBF–funded project that consists of four project partners: Deutsches Klimarechenzentrum (DKRZ), Technische Informationsbibliothek (TIB) and the Universities of Leipzig (Ulei) and Hamburg (UHH).

The objective of the AtMoDat project is to enhance the overall reusability and reuse of atmospheric model data. FAIR[1] data publication standards are well established and implemented in large, internationally coordinated model intercomparison projects (e.g. for climate models in CMIP[2]). These standards, however, cannot be directly transferred to many atmospheric modelling sub-disciplines, such as e.g. urban climate or cloud-resolving modelling. Here, joint standards are still lacking. This is where the AtMoDat project comes into play. Building on the CMIP standard, a new metadata standard for netCDF files was developed which specifically addresses the data standard requirements of atmospheric modelling sub-disciplines: the ATMODAT standard (Ganske et al. 2021).

The AtMoDat project activities are centred around science support, not on generating new scientific data or findings. Within this reporting period, pre-existing model data have been post-processed in order to make them compliant with the ATMODAT standard. Within the coming months, the first ATMODAT compliant datasets will be published at the World Data Center Climate (WDCC) and will serve as a prototype to demonstrate the advantages of an ATMODAT standardization.

Firstly, model output from a small cloud feedback model intercomparison project (cMIP) that has been established under the umbrella of the AEROCOM initiative (aerocom.met.no) is being processed. As suggested by Mülmenstädt et al. (2015), the idea behind this intercomparison project is to use the warm rain frequency as a metric for assessing aerosol-cloud interactions in aerosol-climate models. In total, model data of five cMIP experiments shall be formatted following the ATMODAT standard and subsequently be published at the World Data Center Climate (WDCC). Mainly responsible for this task is ULei whilst the AtMoDat team from DKRZ plays an important advisory role.

Secondly, model data from the microscale obstacle resolving model MITRAS are post-processed towards compliance with the ATMODAT standard. The MITRAS model data are part of a simulation conducted over an approx. 2 km x 2 km area in the city centre of Hamburg. The simulation has a horizontal resolution of 2.5 m; it resolves the urban canopy layer (approx. 220 m in height) with a vertical resolution of 5 m. At this resolution, the complex flow within the heterogeneous urban structures and its interaction with air pollutants

[1] Findable, Accessible, Interoperable, Reusable
[2] Coupled Model Intercomparison Project

can be captured. UHH is mainly responsible for creating and post-processing the data. Since data standardisation in general and, in particular, the self-describing netCDF file format, is only marginally established in this discipline, DKRZ is providing fundamental support in form of expertise and extensive practical advice. Among others, DKRZ has elaborated a variety of new CF standard names for common MITRAS output variables.

Collaboratively, DKRZ and Ulei developed a Python tool to support data producers in checking their data for compliance with the ATMODAT standard. The atmodat data checker is available at https://github.com/AtMoDat/atmodat_data_checker. It will be centrally installed on mistral and will be used by DKRZ data management personal for the general quality control of data prior to their publication via the WDCC.

During the reporting period, the AtMoDat team has performed ample dissemination activities through conference presentations, user workshops and publications, aside from software development and conceptual work. The reference section provides a list of selected dissemination activities.

Originally planned resource utilisation
For the reporting period, it was planned that post-processing a single cMIP experiment would require 336 GB per model year for data storage. It was further assumed that the full model simulation time period was to be post-processed, which is 30 years. Finally, it was planned that, in total, model output from three experiments with ECHAM model variants and output from two experiments with other models (ACME, CAM) were to be post-processed.

The planned space requirements for parallel processing was thus 336 GB x 30 years x 5 models = 50.4 TB on /work; the same space requirements were also estimated for on /arch.

Actual resource utilisation
The actual resource utilisation was substantially lower than originally planned. There are two main reasons. On the one hand, there was a substantial delay in finishing the ATMODAT standard document, which was first released in December 2020. On the other hand, it took longer than expected to develop post-processing routines that convert the model output data into an ATMODAT compliant format. While routines for the cMIP model output have recently been completed and are now being applied to create ATMODAT compliant cMIP datasets, the development of post-processing routines for MITRAS output is still in its beginning. The AtMoDat team is currently preparing the steps towards publishing a first cMIP experiment as ATMODAT compliant dataset with a Digital Object Identifier (DOI) at the World Data Center for Climate (WDCC). This ensures permanent access to the data.

Of the granted 100 node hours on Mistral, only 1 node hour was used so far. So far, the development and testing of the prototype post-processing routines were mainly done on a local ULei PC. In near future, the post-processing of the cMIP experiments and the subsequent automated quality check of the post-processed files will be done on Mistral computing nodes.

Of the granted 49.8 TB space on Lustre work, only 8.2 TB were used so far. More space usage is, however, expected in due term, when the additional cMIP experiments are being transferred to Lustre work for the ATMODAT processing. None of the granted 53.6 TB on

HPSS arch (GB) was used. In the coming allocation period, however, we expect to require around 25 TB for archiving the ATMODAT compliant cMIP experiments.

References

The ATMODAT standard

Ganske, Anette, Kraft, Angelina, Kaiser, Amandine, Heydebreck, Daniel, Lammert, Andrea, Hoeck, Heinke, Thiemann, Hannes, Voss, Vivien, Grawe, David, Leitl, Bernd, Schlünzen, K. Heinke, Kretzschmar, Jan, and Quaas, Johannes: ATMODAT Standard (v3.0), https://doi.org/10.35095/WDCC/ATMODAT_STANDARD_EN_V3_0, 2021.

Peer-reviewed publications

Ganske, A., Heydebreck, D., Höck, H., Kraft, A., Quaas, J., and Kaiser, A.: A short guide to increase FAIRness of atmospheric model data, Meteorologische Zeitschrift, 29, 483–491, https://doi.org/10.1127/metz/2020/1042, 2020.

Ganske, A., Voss, V., Kaiser, A., Heil, A., and Lammert, A.: Die Veröffentlichung von Standardisierten Daten aus der Stadtkli-maforschung, E-Science-Tage 2021: Share Your Research Data. 4. bis 5. März 2021, Online-Konferenz, Veröffentlichung im erscheindenden Tagungsband, heiBOOKs, submitted April 2021, accepted in August 2021.

Conference contributions

Heil, A., Ganske, A., Lammert, A., Heydebreck, D., and Thiemann, H.: The ATMODAT Standard enhances FAIRness of Atmospheric Model data, EMS Annual Meeting 2021, online, 6–10 Sep 2021, EMS2021-298, https://doi.org/10.5194/ems2021-298, 2021

Voss, V., Grawe, D., and Schlünzen, K. H.: How to develop and apply a model data standard on microscale model data., EMS Annual Meeting 2021, online, 6–10 Sep 2021, EMS2021-420, https://doi.org/10.5194/ems2021-420, 2021.

Lammert, A., Ganske, A., Kaiser, A., and Kraft, A.: A Standard for the FAIR publication of Atmospheric Model Data developed by the AtMoDat Project, pico, https://doi.org/10.5194/egusphere-egu21-8144, 2021.

Ganske, A., Kaiser, A., Kraft, A., Heydebreck, D., Lammert, A., and Thiemann, H.: EASYDAB (Earth System Data Branding) for FAIR and Open Data, pico, https://doi.org/10.5194/egusphere-egu21-2139, 2021.

Heydebreck, D., Kaiser, A., Ganske, A., Kraft, A., Schluenzen, H., and Voss, V.: The ATMODAT Standard enhances FAIRness of Atmospheric Model data, Atmospheric Sciences, https://doi.org/10.1002/essoar.10504946.1, 2020.

Kaiser, A., Heydebreck, D., Ganske, A., and Kraft, A.: Making the Maturity of Data and Metadata Visible with Datacite DOIs, Atmospheric Sciences, https://doi.org/10.1002/essoar.10504947.1, 2020.

Voss, V., Schlünzen, K. H., Grawe, D., Heydebreck, D., and Ganske, A.: First results of a model user survey on a mi-cro-scale model data standard, Atmospheric Sciences, https://doi.org/10.1002/essoar.10505067.1, 2020.

Ganske, A., Kaiser, A., and Kraft, A.: Warum und wie Sie Klimamodelldaten veröffentlichen sollten, https://doi.org/10.5194/dkt-12-7, 2020.

Voss, V., Schlünzen, K.H. and Grawe, D.: Entwicklung eines Datenstandards für mikroskalige Modellergebnisse,https://doi.org/10.5194/dkt-12-31, 2020.

Karsten, P., Neumann, D., and Thiemann, H.: Towards increasing the reusability of atmospheric model data: adapt-ing metadata standards and introducing quality criteria.doi:10.5281/zenodo.3667635, 2020.