DeepESDL – an open platform for research and collaboration in Earth Sciences

Gunnar Brandt[1], **Alicja Balfanz**[1], Norman Fomferra[1], Tejas Morbagal Harish[1], Miguel Mahecha[2], Guido Kraemer[2], David Montero[2], Stephan Meißl[3], Stefan Achtsnit[3], Josefine Umlauft[4], Anja Neumann[4], Alex Horton[5], Martin Ewart[5], Fabian Gans[6], and Anca Anghelea[7]

[1]Brockmann Consult GmbH, Hamburg, Germany

[2]Remote Sensing Centre for Earth System Research, University of Leipzig, Leipzig, Germany

[3]EOX IT Services GmbH, Vienna, Austria

[4]Center for Scalable Data Analytics and Artificial Intelligence, University of Leipzig, Leipzig, Germany

[5]Earthwave Ltd., Edinburgh, United Kingdom

[6]Department Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany

[7]ESA-ESRIN, Frascati, Italy

The Deep Earth System Data Lab (DeepESDL, https://earthsystemdatalab.net) provides an AI-ready, collaborative environment enabling researchers to understand the complex dynamics of the Earth System using numerous datasets and multi-variate, empirical approaches. The solution builds on work done in previous projects funded by the European Space Agency (CAB-LAB and ESDL), which established the technical foundations and created measurable value for the scientific community, e.g., Mahecha et al. (2020, https://doi.org/10.5194/esd-11-201-2020) or Flach et al. (2018, https://doi.org/10.5194/bg-15-6067-2018). DeepESDL relies heavily on the well-established open-source technology stacks for data science in Python, thus ensuring usability and compatibility.

The core of the DeepESDL is represented by the provision of programmatic access to various data sources in analysis-ready form, organised in data cubes combined with adequate computational resources and capabilities to allow researchers to immediately focus on efficient analysis and of multi-variate and high-dimensional data through empirical methods or AI approaches.

To ensure proper documentation and discoverability, DeepESDL is building an informative catalogue to find all available data and to find the required metainformation describing them. This includes not only standard information, e.g., regarding spatial and temporal coverage, versioning, but also on specific transformation methods applied during data cube generation.

The system design has openness, collaboration, and dissemination as key guiding principles. As science teams need proper tooling support to efficiently work together in this virtual environment, one of the key elements of the architecture is represented by the DeepESDL Hub, providing teams of scientific users with the means for collaboration and exchange of versioned results, source codes, models, execution parameters, and other artifacts and outcomes of their activities in a

simple, safe and reliable way. The tools are complemented by an integrated, state-of-the-art application for the visualisation of all data in the virtual laboratory including input data, intermediate results, as well as the final products.

Furthermore, the DeepESDL supports the implementation and execution of Machine Learning workflows on Analysis Ready Data Cubes in a reproducible and FAIR way, allowing sharing and versioning of all ML artifacts like code, data, models, execution parameters, metrics, and results as well as tracking each step in the ML workflows (supported by integration with Open-Source tools like TensorBoard or Mlflow) for an experiment so that others can reproduce them and contribute.

Finally, dissemination is essential for the Open Science spirit of the DeepESDL. Two applications, xcube Viewer and 4D viewer, offer comprehensive user interfaces for interactive exploration of multi-variate data cubes. Both use the same RESTful data service API provided by xcube Server. The latter also provides OGC interfaces, so that other OGC-compliant applications, such as QGIS3, are able to visualise analysis-ready data cubes generated within DeepESDL.

To foster collaboration, additional features such as publishing individual Jupyter Notebooks as storytelling documents or even books using Jupyter Books or the Executable Book Project are being explored, together with concepts such as storytelling and DeepESDL User Project Dashboards which may also link to the viewers and Notebooks.