

EGU24-20127, updated on 13 Mar 2024

<https://doi.org/10.5194/egusphere-egu24-20127>

EGU General Assembly 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



An interlinked research data infrastructure for time-series data from the Helmholtz Research Field Earth & Environment

Christof Lorenz¹, Benjamin Louisot¹, Sabine Barthlott², Benjamin Ertl², Linda Baldewein³, Ulrike Kleeberg³, Marie Ryan³, Nils Brinckmann⁴, Marc Hanisch⁴, Roland Koppe⁵, Marc Adolf⁶, Claas Faber⁶, Andreas Lehmann⁶, David Schäfer⁷, Ralf Kunkel⁸, Ulrich Loup⁸, Jürgen Sorg⁸, and Hylke van der Schaaf⁹

¹Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany

²Institute of Meteorology and Climate Research - Atmospheric Trace Gases and Remote Sensing (IMK-ASF), Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany

³Helmholtz Coastal Data Center (HCDC), Helmholtz-Zentrum Hereon, Geesthacht, Germany

⁴Section 5.2 - eScience Centre, Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany

⁵Computing and Data Centre - Data Division, Alfred Wegener Institute - Helmholtz Centre for Polar and Marine Research (AWI), Bremerhafen, Germany

⁶Information, Data and Computing Centre, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

⁷Research Unit Smart Models and Monitoring - Department of Monitoring and Exploration Technologies, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

⁸Institute of Bio- and Geosciences - Agrosphere (IBG-3), Jülich Research Centre (FZJ), Jülich, Germany

⁹Information Management and Production Control, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Karlsruhe, Germany

Time-series data are crucial sources of reference information in all environmental sciences. And beyond typical research applications, the consistent and timely publication of such data is increasingly important for monitoring and issuing warnings, especially in times of growing frequencies of climatic extreme events. In this context, the seven Centres from the Helmholtz Research Field Earth and Environment (E&E) operate some of the largest environmental measurement-infrastructures worldwide. These infrastructures range from terrestrial observation systems in the TERENO observatories and ship-borne sensors to airborne and space-based systems, such as those integrated into the IAGOS infrastructures.

In order to streamline and standardize the usage of the huge amount of data from these infrastructures, the seven Centres have jointly initiated the STAMPLATE project. This initiative aims to adopt the Open Geospatial Consortium (OGC) SensorThings API (STA) as a consistent and modern interface tailored for time-series data. We evaluate STA for representative use-cases from environmental sciences and enhance the core data model with additional crucial metadata such as data quality, data provenance and extended sensor metadata. After centre-wide implementation, the standardized STA interface also serves community-based tools, e.g., for data visualization, data access, quality assurance/quality control (QA/QC), or the management of monitoring systems. By

connecting the different STA endpoints of the participating research Centres, we establish an interlinked research data infrastructure (RDI) and a digital ecosystem around the OGC SensorThings API tailored towards environmental time-series data.

In this presentation, we want to show the status of the project and give an overview of the current data inventory as well as linked tools and services. We will further demonstrate the practical application of our STA-based framework with simple and representative showcases. With our contribution, we want to promote STA for similar applications and communities beyond our research field. Ultimately, our goal is to provide an important building block towards fostering a more open, FAIR (Findable, Accessible, Interoperable, and Reusable), and harmonized research data landscape in the field of environmental sciences.