



## The STAMPLATE-Schema as a unifying metadata language for FAIR and AI-ready environmental time-series data in the DataHub ecosystem

**Christof Lorenz**<sup>1</sup>, Nils Brinckmann<sup>2</sup>, Claas Faber<sup>3</sup>, Marc Hanisch<sup>2</sup>, Roland Koppe<sup>4</sup>, Ralf Kunkel<sup>5</sup>, Ulrich Loup<sup>5</sup>, Mihir Rambhia<sup>6</sup>, Hylke van der Schaaf<sup>7</sup>, David Schäfer<sup>8</sup>, and the STAMPLATE-Team within the DataHub Initiative of the Research Field Earth and Environment\*

<sup>1</sup>Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe & Garmisch-Partenkirchen, Germany

<sup>2</sup>Secion 5.2 - eScience Centre, GFZ Helmholtz Centre for Geosciences, Potsdam, Germany

<sup>3</sup>Ocean Research Technology Centre, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

<sup>4</sup>Software Engineering, Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven, Germany

<sup>5</sup>Institute of Bio- and Geosciences - Agrosphere (IBG-3), Jülich Research Centre (FZJ), Jülich, Germany

<sup>6</sup>Helmholtz Coastal Data Center (HCDC), Helmholtz-Zentrum Hereon, Geesthacht, Germany

<sup>7</sup>Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany

<sup>8</sup>Department of Monitoring and Exploration Technologies - Research Data Management (RDM), Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

\*A full list of authors appears at the end of the abstract

Environmental observations from sensor systems remain one of the most important sources of ground-truth data in Earth System Sciences. In particular, the rapid rise of AI-based methods, high-resolution modelling, and the growing demand for near-real-time reference data to support environmental decision-making, have substantially increased the need for reliable, interoperable, and AI-ready observational data.

To enable seamless integration and effective use of such data across diverse application scenarios (especially when combining observations from multiple sources), consistent data structures, well-defined interfaces, and harmonised, machine-readable metadata are essential. These requirements represent both a technical and a community-driven challenge and form a key prerequisite for ensuring the AI-readiness of sensor data.

Within the Helmholtz Research Field *Earth & Environment*, the DataHub initiative addresses this challenge by developing a uniform and FAIR research data infrastructure for observational time-series data across all seven contributing German Helmholtz Centres. Central to this infrastructure is the OGC SensorThings API STAMPLATE-Schema, a unified metadata schema for sensor-based observational data. The STAMPLATE-Schema serves as the semantic backbone of the DataHub ecosystem, providing a shared, machine-actionable language to describe deployments, sensors and observations. It is built upon JSON-LD and schema.org, enabling semantic interoperability, extensibility, and direct compatibility with web technologies and AI workflows.

The STAMPLATE-Schema connects and aligns the core ecosystem components, including the Sensor Management System (SMS) – which provides user-friendly management of sensor and deployment metadata - and the Earth Data Portal (EDP), which supports cataloguing, discovery, and visualisation of SensorThings API-based data. Additional integrations, such as the System for automated Quality Control (SaQC) and the time-series handling via time.io, build on this shared metadata foundation and support typical observational data workflows including data flagging, quality assessment, and downstream processing.

The STAMPLATE-Schema and the associated federated SensorThings API-based data infrastructures are currently being implemented across several major German research centres and large-scale observational projects, including the TERENO-network with its multiple observatories. Together, they are expected to provide access to more than 20 billion observations from seven research centres spanning multiple environmental research domains, including terrestrial, atmospheric, and marine systems, by the end of the year.

The DataHub and the STAMPLATE-Schema thus provide a common metadata language and framework for FAIR and AI-ready sensor data across our research field and similar federated research data infrastructures.

**STAMPLATE-Team within the DataHub Initiative of the Research Field Earth and**

**Environment:** Martin Abbrent, Marc Adolf, Sabine Barthlott, Nils Brinckmann, Jan Bumberger, Kirsten Elger, Claas Faber, Marc Hanisch, Ulrike Kleeberg, Roland Koppe, Ralf Kunkel, Christof Lorenz, Benjamin Louisot, Ulrich Loup, Florian Obersteiner, Mihir Rambhia, Marie Ryan, Jasper Schalla, David Schäfer, Thomas Schnicke, Emanuel Söding, Jürgen Sorg