Beitrag ID: **17**                                             Typ: **Presentation**

# DS2STAC: A Python package for harvesting and ingesting (meta)data into STAC-based catalog infrastructures

*Freitag, 9. Juni 2023 09:45 (15 Minuten)*

Despite the vast growth in accessible data from environmental sciences over the last decades, it remains difficult to make this data openly available according to the FAIR principles. A crucial requirement for this is the provision of metadata through standard catalog interfaces or data portals for indexing, searching, and exploring the stored data.

With the release of the community-driven Spatio-Temporal Assets Catalog (STAC), this process has been substantially simplified as STAC is based on highly flexible and lightweight GeoJSONs instead of large XML-files. The number of STAC-users has hence rapidly increased and STAC now features a comprehensive ecosystem with numerous extensions. This is also why we have chosen STAC as our central catalog framework in our research project Cat4KIT, in which we develop an open-source software stack for the FAIRification of environmental research data.

A central element of this project is the automatic (meta)data and service harvesting from different data servers, providers and services. This so-called DS2STAC-module hence contains tailormade harvesters for different data sources and services, a metadata validator and a database for storing the STAC items, collections and catalogs. Currently, DS2STAC can be used for harvesting from THREDDS-Server, Intake-Catalogs, and SensorThings APIs. In all three cases, it creates and manages consistent STAC-items, -catalogs and -collections which are then made openly available through the pgSTAC-database and the STAC-FastAPI to allow for a user-friendly interaction with our environmental research data.

In this presentation, we want to demonstrate DS2STAC and also show its functionalities within our Cat4KIT-framework. We also want to discuss further use-cases and scenarios and hence propose DS2STAC as a modular tool for harvesting (meta)data into STAC-based catalog infrastructures.

**Hauptautor:**   HADIZADEH, Mostafa (Karlsruher Institut für Technologie - Institut für Meteorologie und Klimaforschung Atmosphärische Umweltforschung (IMK-IFU), KIT-Campus Alpin)

**Co-Autoren:**   Dr. LORENZ, Christof (Karlsruher Institut für Technologie - Institut für Meteorologie und Klimaforschung Atmosphärische Umweltforschung (IMK-IFU), KIT-Campus Alpin); Dr. BARTHLOTT, Sabine (Karlsruher Institut für Technologie - Institut für Meteorologie und Klimaforschung, Department Atmosphärische Spurengase und Fernerkundung (IMK-ASF)); Dr. FÖSIG, Romy (Karlsruher Institut für Technologie - Institut für Meteorologie und Klimaforschung, Department Atmosphärische Aerosol Forschun (IMK-AAF)); Dr. ÇAYOĞLU, Uğur (Karlsruher Institut für Technologie - Steinbuch Centre for Computing (SCC)); Dr. ULRICH, Robert (Karlsruher Institut für Technologie - Bibliothek (BIB)); Dr. BACH, Felix (FIZ Karlsruhe –Leibniz-Institut für Informationsinfrastruktur)

**Vortragende(r):**   HADIZADEH, Mostafa (Karlsruher Institut für Technologie - Institut für Meteorologie und Klimaforschung Atmosphärische Umweltforschung (IMK-IFU), KIT-Campus Alpin)

**Sitzung Einordnung:**  Talks