Contribution ID: **30** Type: **Poster**

An Interactive FAIR-Data Publishing System for Time-Series Data

Tuesday 10 October 2023 14:30 (15 minutes)

An important point for the widespread dissemination of FAIR-data is the lowest possible entry barrier for preparing and providing data to other scientists according to the FAIR criteria. If scientists have to manually extract, transform and annotate the data according to the FAIRcriteria and then export it to make it available to the public, this requires a significant investment of time that does not primarily reward the scientists who prepares and provides the data.
The Energy-Lab at KIT is running a large cluster of an Influx database management system in which energy related time-series data being stored in a variety of individual databases for over 10 years. In order to increase the willingness to make data available to the scientific public, we have developed a web application that greatly supports and automates the publication and annotation process of time-series-data stored in Influx databases.
The web application consists of a backend service and an interactive frontend. The backend provides arbitrary, predefined and annotated time-series-data of a measurement (an Influx database structure that corresponds to a table in a relational database) via an URL without requiring any further information for access. The specification of the data is done via HTTP-GET parameters. These include the desired time interval and specific conditions on the attributes as well as a configuration file in which the Influx-server access information are stored. The actual request is made by a series of REST API calls to the InfuxDB. In order to be able to extract arbitrarily large amounts of data, a stream-based approach was chosen. The data is returned as an RO-Crate dataset, using the CSVW Namespace to describe tabular data. The column data-types are extracted from metadata calls to the Influx database. Further information about the attributes (like unit, quantity) can be specified in the configuration file.
The frontend implements the interactive construction of the URL for reading out the time-series-data. The first step is to select the specific configuration file stored for a particular measurement, which contains the information (see above) for accessing a specific database. This information is used, to access the measurement and determine the time interval for which data is available. Meta information of the measurement is read out including the attributes with their data types. In addition, for attributes which act as tags (descriptive attributes), the existing tag-values are extracted. These attributes can be used to interactively formulate extraction conditions (e.g. only data of certain buildings, devices,...). Finally, the time interval of the data to be extracted must be specified. The result of this step is a URL, conforming to the backend API, to export the data in the appropriate format. Currently all attributes (tags & fields) are returned.
In a future version we also plan that it is possible to specify which attributes should be returned. Further extensions include a cache mechanism, where only dynamic data must be retrieved on every request while historical data requests can be fulfilled from a cache.

Please assign your contribution to one of the following topics

Technological solutions for findable and machine-readable metadata

Please specify ”other” (stakeholder)

In addition please add keywords.

Data-exporter, CO-crate, Export-configurator, time-series data

Please assign yourself (presenting author) to one of the stakeholders.

Researchers

Primary author: SCHMIDT, Andreas (KIT)

Co-authors: Mr KOUBAA, Mohamed Anis (Institute for Automation and applied Informatics); SCHWEIKERT, Jan (KIT); STUCKY, Karl-Uwe (KIT); SUESS, Wolfgang (KIT)

Presenter: SCHMIDT, Andreas (KIT)

Session Classification: Poster session

Track Classification: Poster session