

# Link My Data

## Community-Based Curation of Environmental Sensor Data

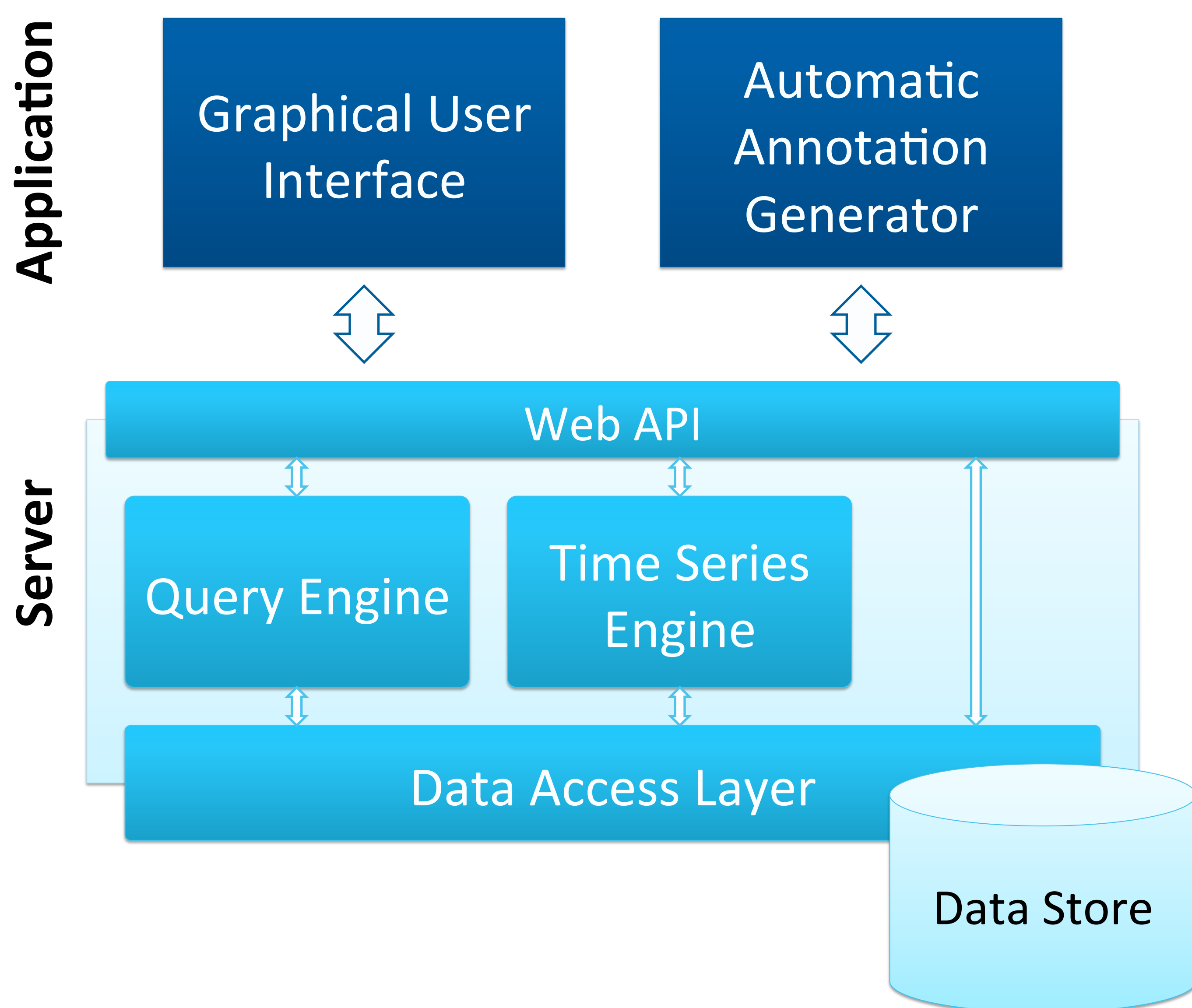
Heiko Müller, Chris Peters, Peter Taylor, Andrew Terhorst

CSIRO COMPUTATIONAL INFORMATICS  
www.csiro.au



One of the biggest obstacles to reuse of third-party sensor data is a lack of knowledge about data properties (e.g., provenance and quality) leading to a lack of trust in the data. **Link My Data (LMD)** is a first step towards overcoming this problem. LMD provides a platform for data curation that allows users to share knowledge about individual sensors and sensor observations. The system supports annotation and transformation of sensor data on the Web to improve data quality and (re-)usability.

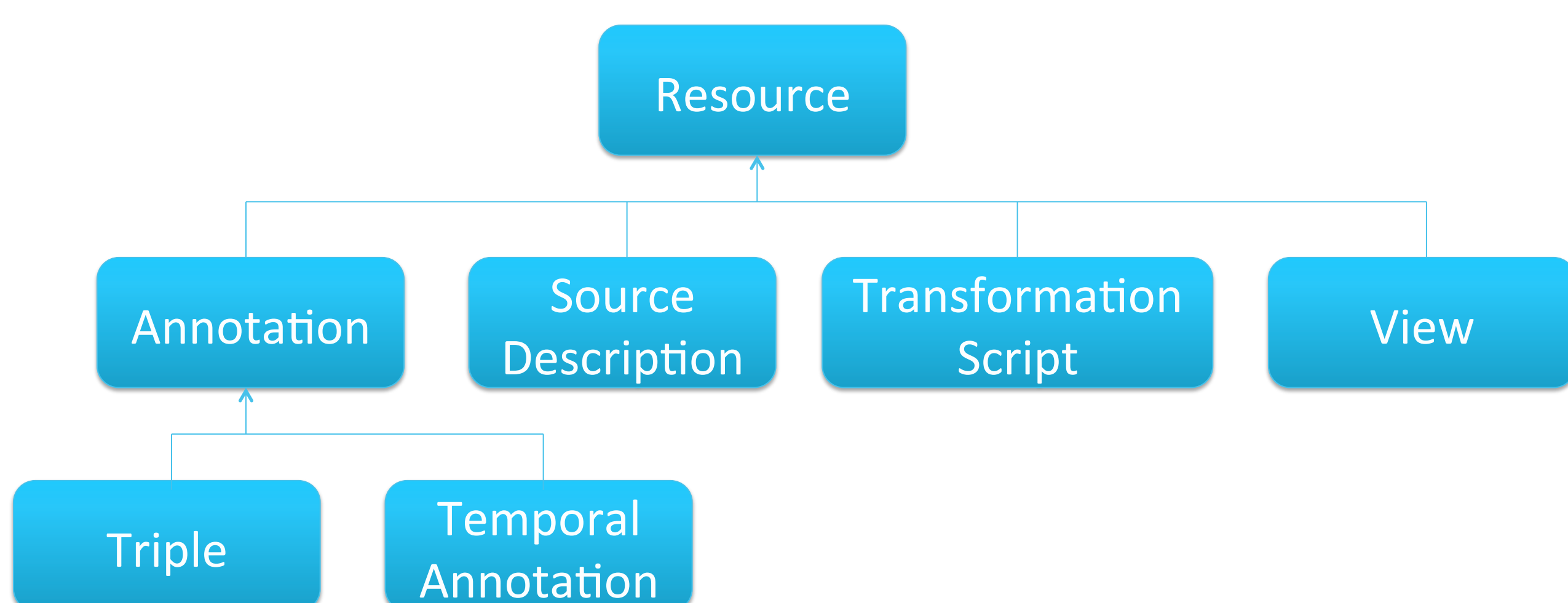
### System Architecture



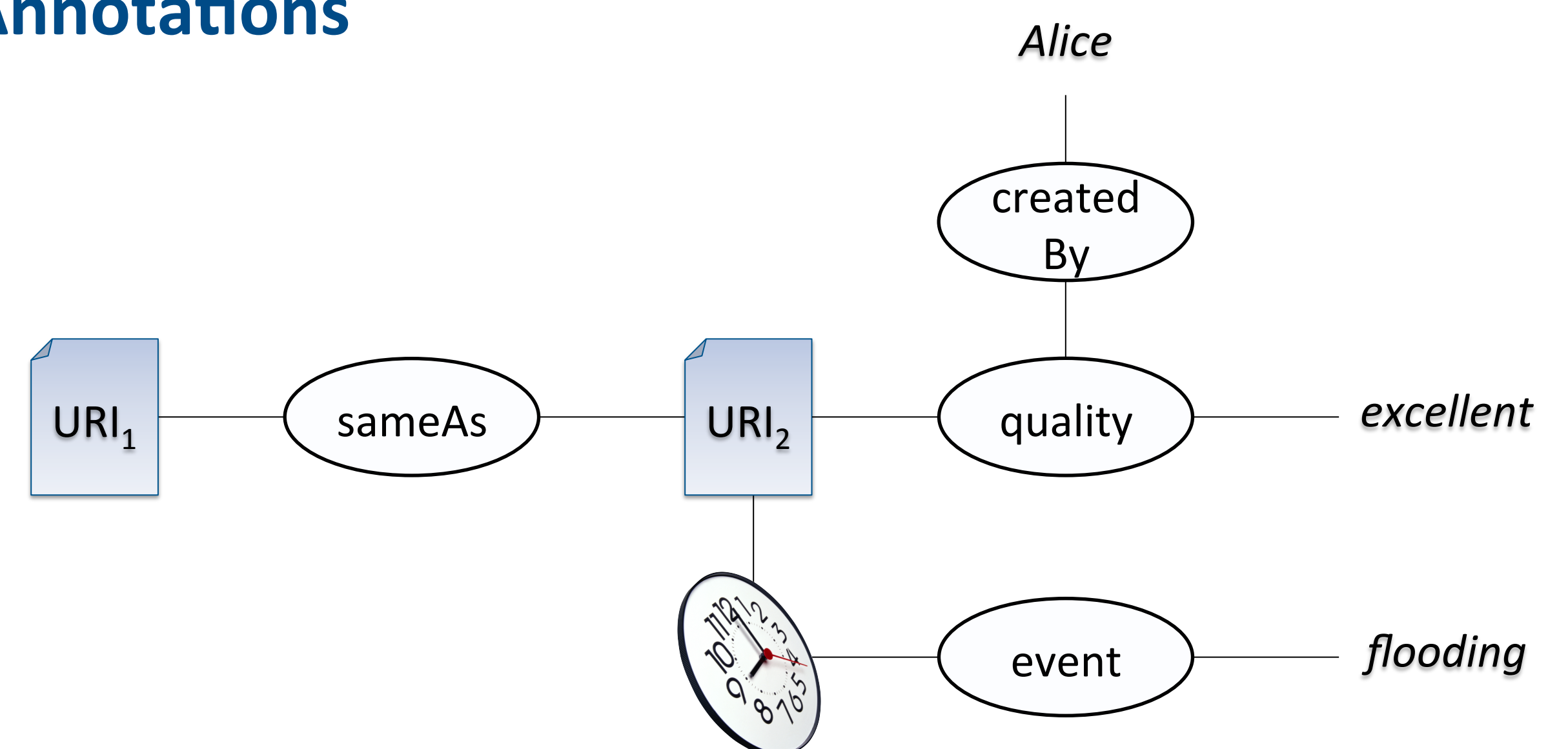
The system is implemented in Java and run as a web service using Apache Tomcat. Data is stored in a dedicated data store. There are four main components in the architecture: The *Data Access Layer* provides data access and abstracts from the underlying data store. The *Query Engine* allows searching for objects with specific annotations (e.g., sensors in a region that observe rainfall in a specified data format). The *Time Series Engine* is responsible for executing data transformation programs. The system is accessible via a *Web API*.

### Data Model

In LMD, any resource with a URI can be annotated. The system itself provides URLs for all system generated objects. There are three types of annotations: literals, links and temporal annotations. Literals and links are similar to triples in RDF. Temporal annotations support definition of intervals and points in time that are of importance to a resource. Given that all annotations have URLs, we can annotate them as well.

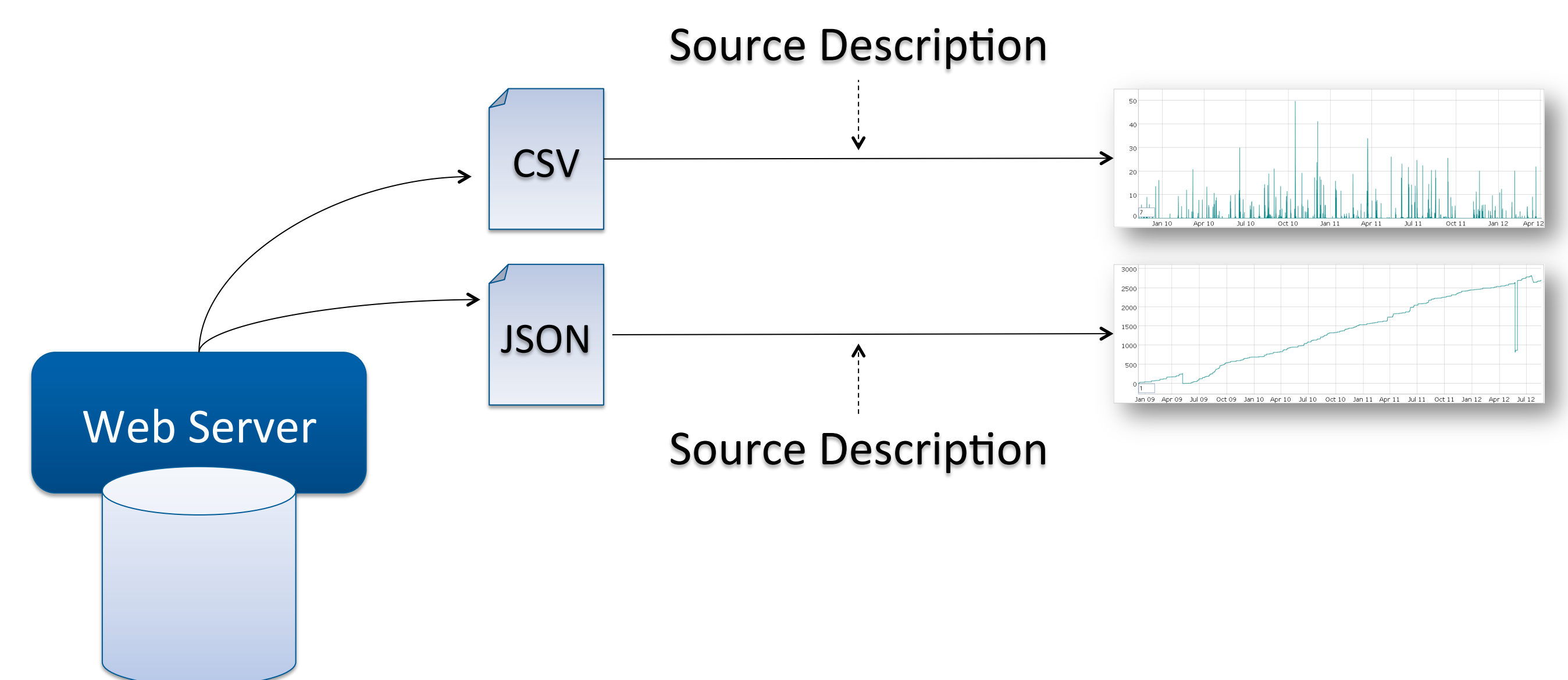


### Annotations



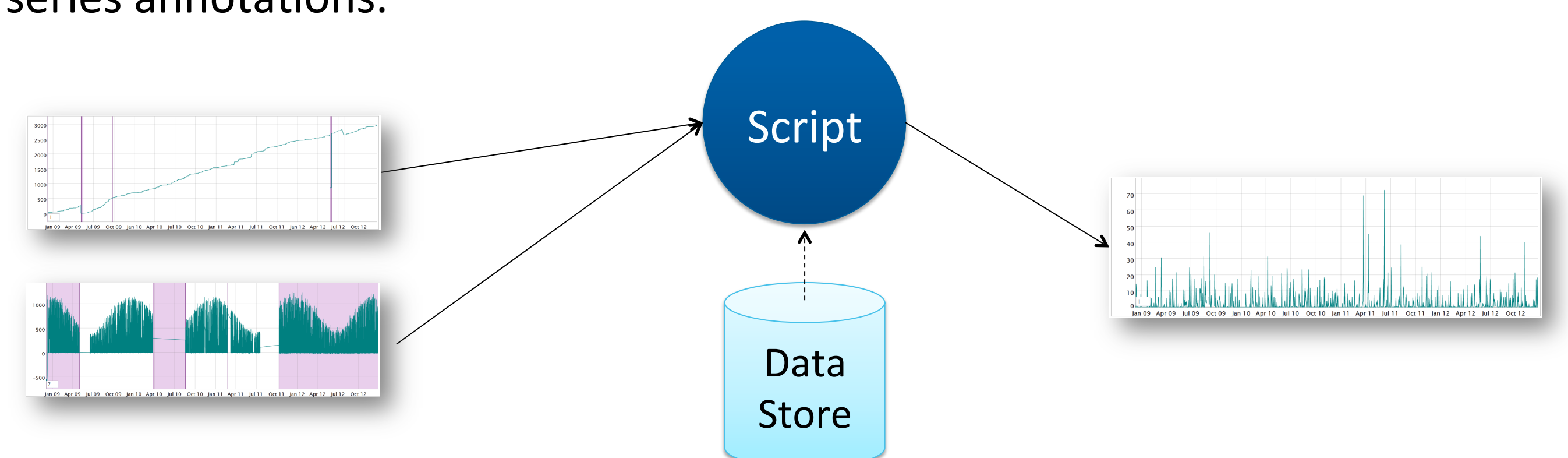
### Source Descriptions

Time series in LMD are sequences of (time, value)-pairs that are published as (part of) structured documents on the Web. Each document may contain multiple time series. LMD allows users to define *source descriptions* that specify how to extract time series from documents.



### Transformation Scripts

LMD is equipped with a set of operators for transforming time series data. This allows users to manipulate existing data and share the results. A special operator allows generation of transformation pipelines on-the-fly from time series annotations.



- **SELECT**: Filter based on date and/or value.
- **UPDATE**: Modify values
- **REPLACE**: Replace parts with values from other time series.
- **GROUP BY**: Aggregate values in temporal intervals.
- **SLIDE WINDOW**: Aggregate values in sliding window.
- **MERGE**: Align and merge two or more time series.
- **FOR EACH**: Use annotations to instantiate transformation operators.

#### FOR FURTHER INFORMATION

Heiko Müller  
Intelligent Sensing and Systems Laboratory  
e heiko.mueller@csiro.au  
W www.ict.csiro.au/staff/heiko.mueller/

#### ACKNOWLEDGEMENTS

The Intelligent Sensing and Systems Laboratory and the Tasmanian node of the Australian Centre for Broadband Innovation is assisted by a grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts.