Digital Object Architecture data layer over a network storage system

DATA TO INSIGHT CENTER
INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

PRAGMA

Yu Luo, Beth Plale, Jeremy Musser, Martin Swany
Indiana University Bloomington, USA

Motivation

Communities and researchers are exploring the application of smart technology to enhance urban and rural communities. These efforts fall under the rubric of smart and connected communities (S&CC). In one case familiar to PRAGMA, Airbox style sensors were deployed in Salt Lake City (SLC) in the US to detect small pockets of PM 2.5 levels in this hilly region. Open Science principles can be applied to build trust and transparency [1], and developing trust is an integral part to smart and connected communities partnerships [2]. A key principle of open science is that data be given a unique persistent ID (PID), such as the DOI. PIDs are a key principle because uniquely identifying data is important for trusting a reference to data in the future.

Research

Research Data Alliance (RDA) community is exploring a global Digital Object Architecture (DOA) [3] for FAIR data (data that is findable, accessible, interoperable, and reusable). FAIR data will reside in databases, in large scale data centers, and in trusted repositories. Data can be universally discoverable through the DOA.

A novel data storage approach is the network storage system, used by applications like Named Data Networking.

We explore how a network storage system (UNIS [4] of Martin Swany) will interoperate with the DOA layer and take advantage of the PID Kernel Information [5] (Beth Plale), and the E-RPID testbed (Rob Quick).
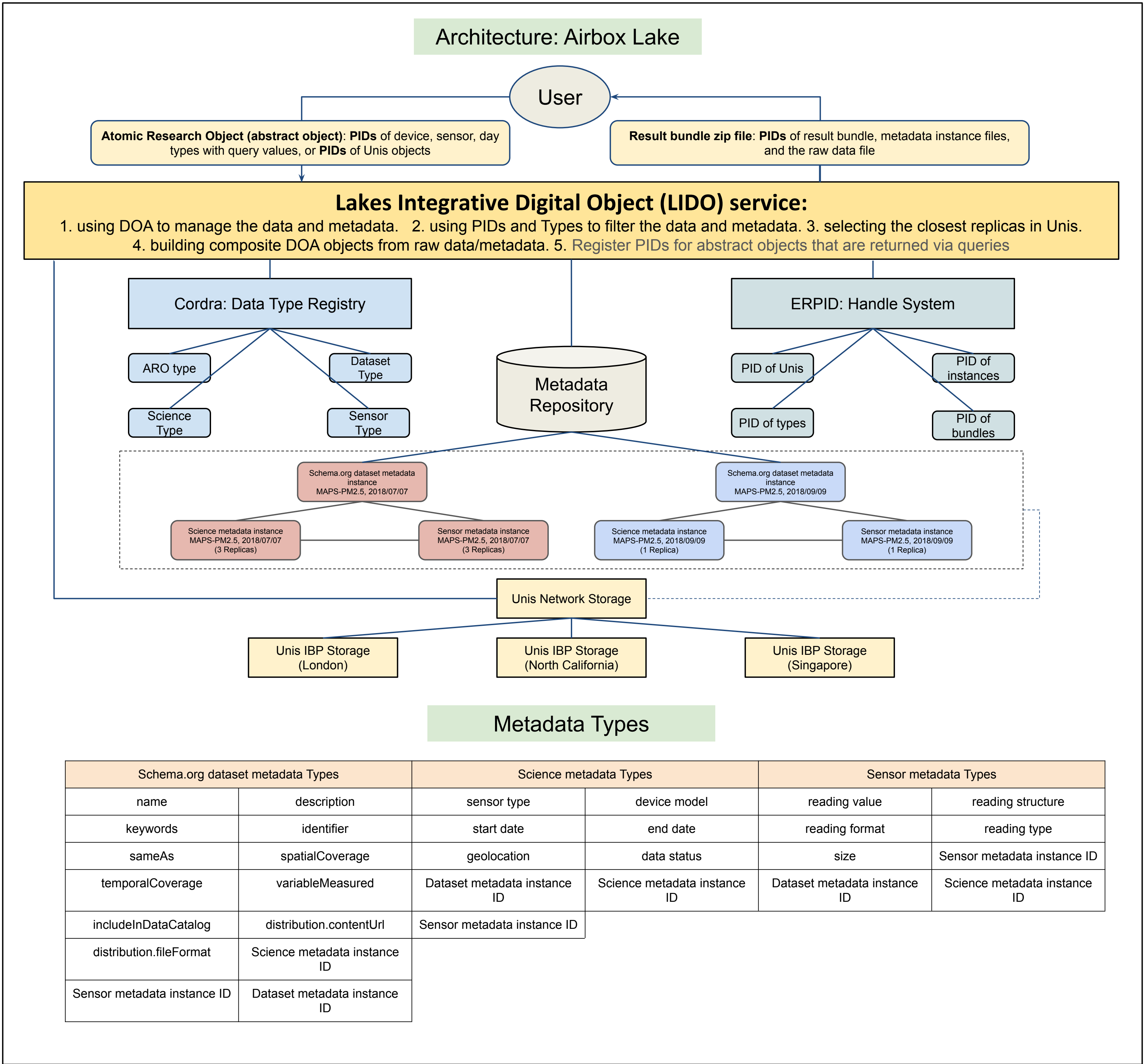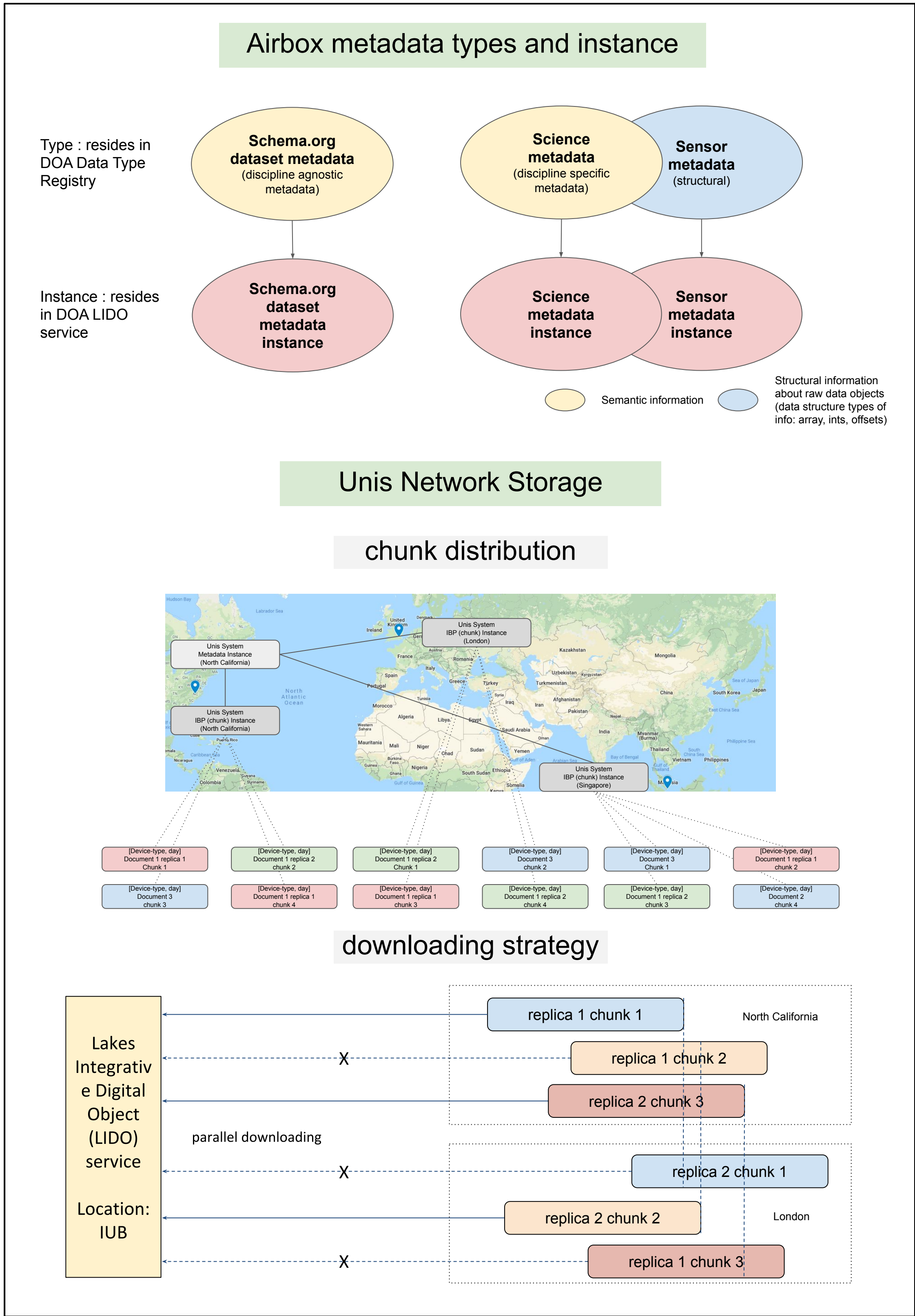
Our research is:

1. developing a **Lakes Integrative Digital Object** (**LIDO**) that delivers to an application a set of logical objects in the DOA space, while managing the raw data in Unis, and
2. LIDO's ability to contribute to trust at the local level.

Architecture and Approach

In our work to date we have set up a testbed using the Airbox data we curated with persistent IDs in a partnership with NCHC. We are:

- Storing the Airbox objects to the Unis Network Storage.
- Creating domain agnostic metadata for the airbox data that uses the Schema.org Dataset schema.
- Defining a set of abstract objects at the DOA layer, and are
- Developing the **Lakes Integrative Digital Object** (**LIDO**) to extract information about the storage of the data in a global network storage

Metadata Types

Schema.org dataset metadata Types		Science metadata Types		Sensor metadata Types	
name	description	sensor type	device model	reading value	reading structure
keywords	identifier	start date	end date	reading format	reading type
sameAs	spatialCoverage	geolocation	data status	size	Sensor metadata instance ID
temporalCoverage	variableMeasured	Dataset metadata instance ID	Science metadata instance ID	Dataset metadata instance ID	Science metadata instance ID
includeInDataCatalog	distribution.contentUrl	Sensor metadata instance ID			
distribution.fileFormat	Science metadata instance ID				
Sensor metadata instance ID	Dataset metadata instance ID				

Acknowledgements

References

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3.

[2] GovEx. First things first: laying the foundation for a smart city, 2018; https://bit.ly/2NdSvlW

[3] Kahn, R., & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries*, 6(2), 115-123.

[4] Portnoi, M., & Swany, M. Unified Network Information Services (UNIS) allow users to discover network services and capabilities. *SIGCOMM Computer Communication Review*, 27, 3-14.

[5] Weigel, Tobias, Plale, Beth, Parsons, Mark, Zhou, Gabriel, Luo, Yu, Schwardmann, Ulrich, & Quick, Robert. (2018, August 9). RDA Recommendation on PID Kernel Information (draft). Zenodo. http://doi.org/10.5281/zenodo.1462543