Linked Data in the Built Environment: Standards Projects and Other Interesting Things

James O'Donnell and Cathal Hoare

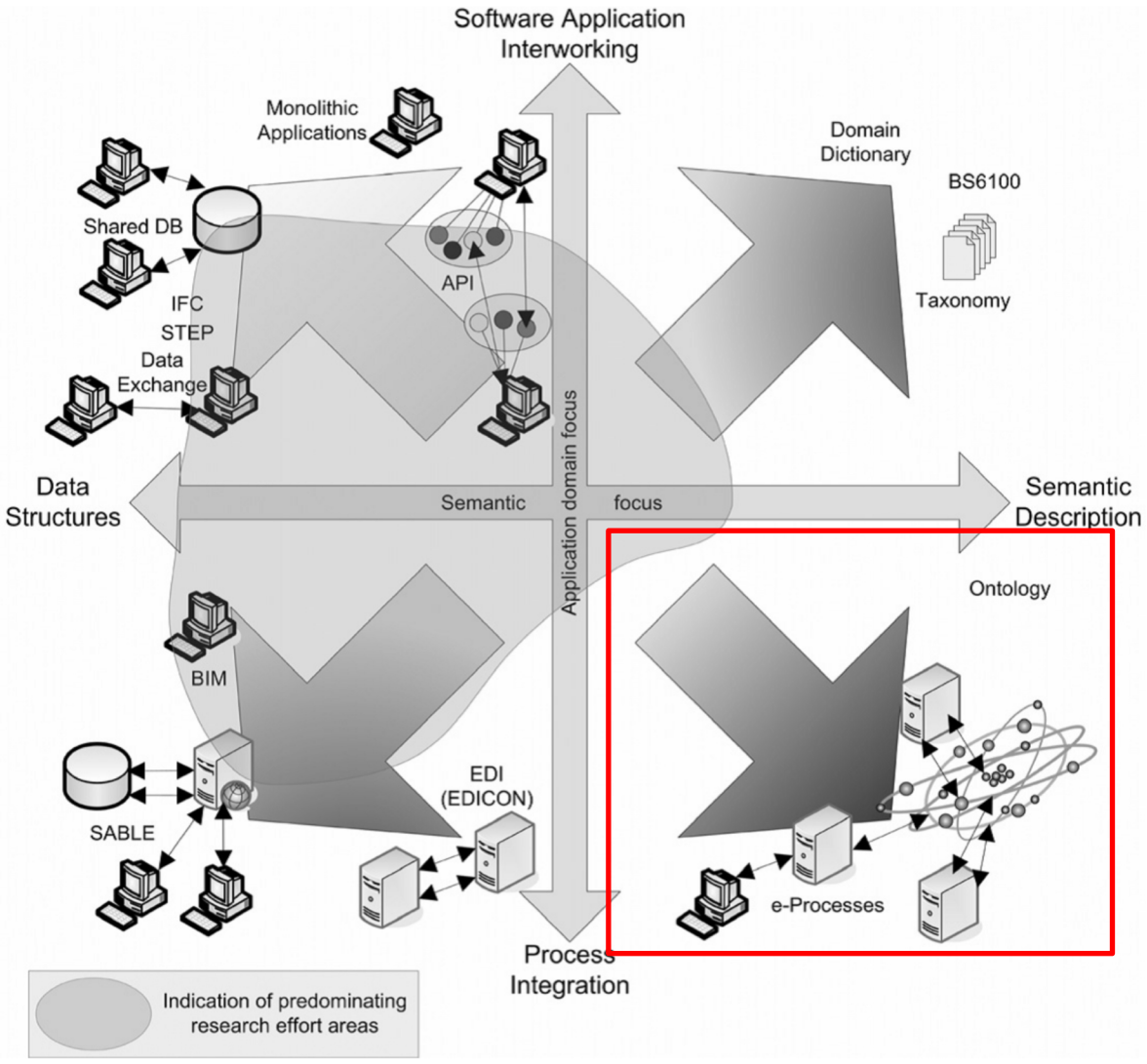
04/4/2019

IBPSA Project 1 Expert Meeting

RWTH Aachen, Germany

General

Only a little research is taking place in the quadrant defined by process integration and semantic description



Semantic web enabled data definition can be achieved with less up-front cost, extremely useful where data changes rapidly.

Area with greatest research potential?

Stefan Boddy, et. al.. Computer integrated construction: A review and proposals for future direction. *Advances in Engineering Software*, 38(10):677–687, October 2007. ISSN 0965-9978. doi: 10.1016/j. advengsoft.2006.10.007. URL http://www.sciencedirect.com/science/article/pii/ S096599780600175X.

Semantic Web technologies and Linked Data provide many benefits that include:

- Improve data interoperability and exchange (though challenges remain);
- The ability to reason about data (multiple data sources);
- Makes for flexible and maintainable software architectures;
- Provide opportunities to solve other issues such as data privacy;

Linked Data Overview

- Related standards are managed by W3C

- Linked data can be seen as a network of semantically organized data repositories.
  - A large number of technologies have matured around semantic standards
  - This allows for architectures to solve some of the afore mentioned problems

- Each repository is identified using a Uniform Resource Identifier (URI).
- Each repository is defined using RDF/OWL and queried using SPARQL.

What is Linked Data

- **R**esource **D**escription **F**ramework (RDF) - identifies a resource with a URI and provides a set of properties that describe the entity. RDF is expressed using XML and can thus be easily exchanged and parsed as input to some computational process.

- Web Ontology Language (OWL) – defines the structure of knowledge for various domains using nouns to describe entities and verbs to describe the relationships between these entities.

- SPARQL – Is a query language that can be used to query and manipulate RDF data stores.

Linked Data in use

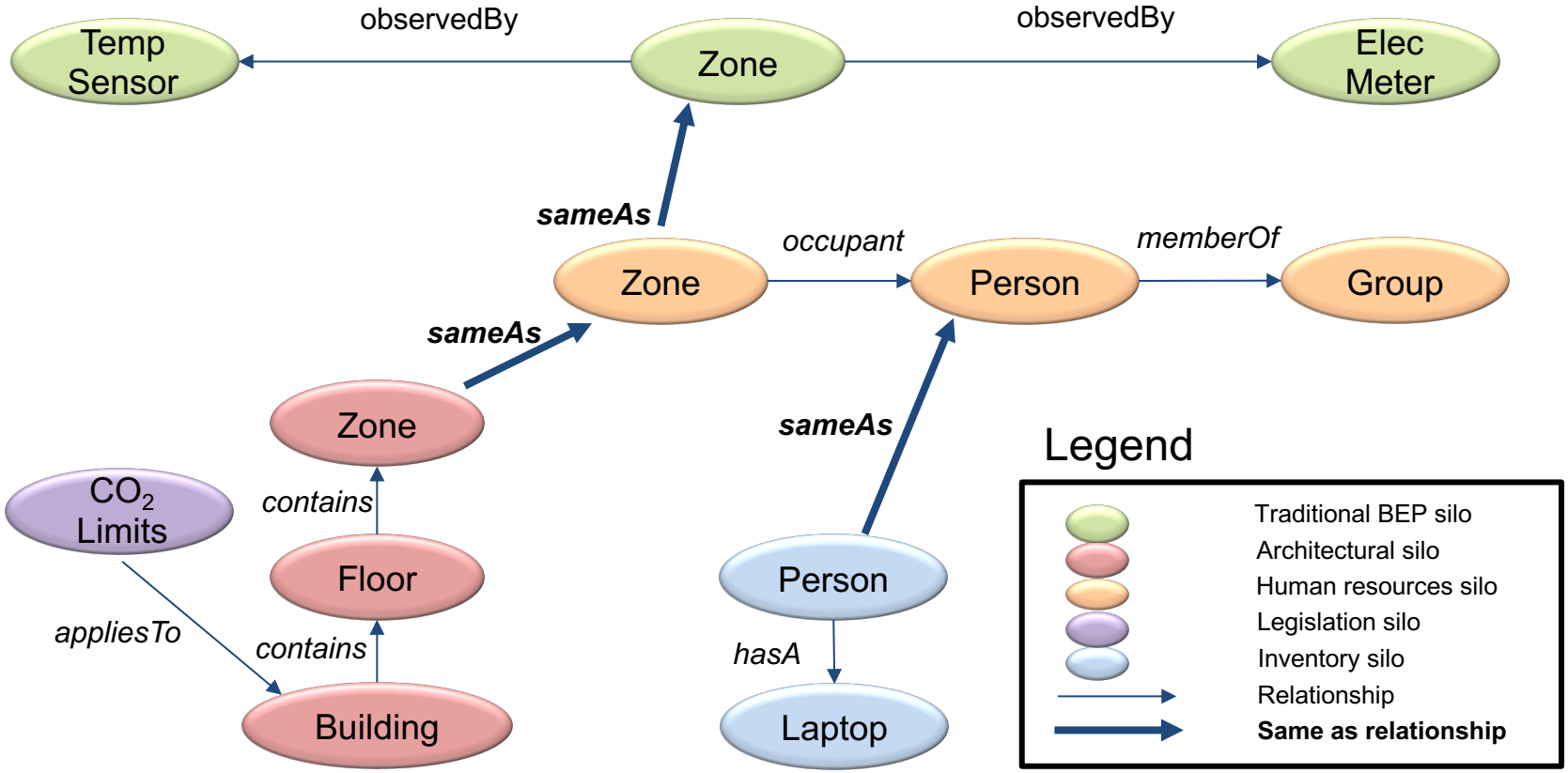
- W3C proposes numerous related standards including:
  - Data Cube - enables a statistical dataset to be organized along a group of dimensions, together with associated metadata.
    - QB4ST provides specific support for spatio-temporal data.
  - DCAT (Data Catalogue) - DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs (information about information).
  - Linked Data Notifications – a protocol that describes how servers can have messages pushed to them by applications and how other applications may retrieve those messages.
- All of these standards allow their data to be reasoned about.

A small selection of Available ontologies

- IFC
- CityGML
- SimModel
- Semantic Sensor Network (SSN)
- Building Topology Ontology (BOT)
- Energy Efficiency Ontology (EEOnt)
- DOCK 1.0
- OnToShare
- SWIMS
- PCO
- OntoSensor
- PV-TONS
- ThinkHome
- Many many more…..

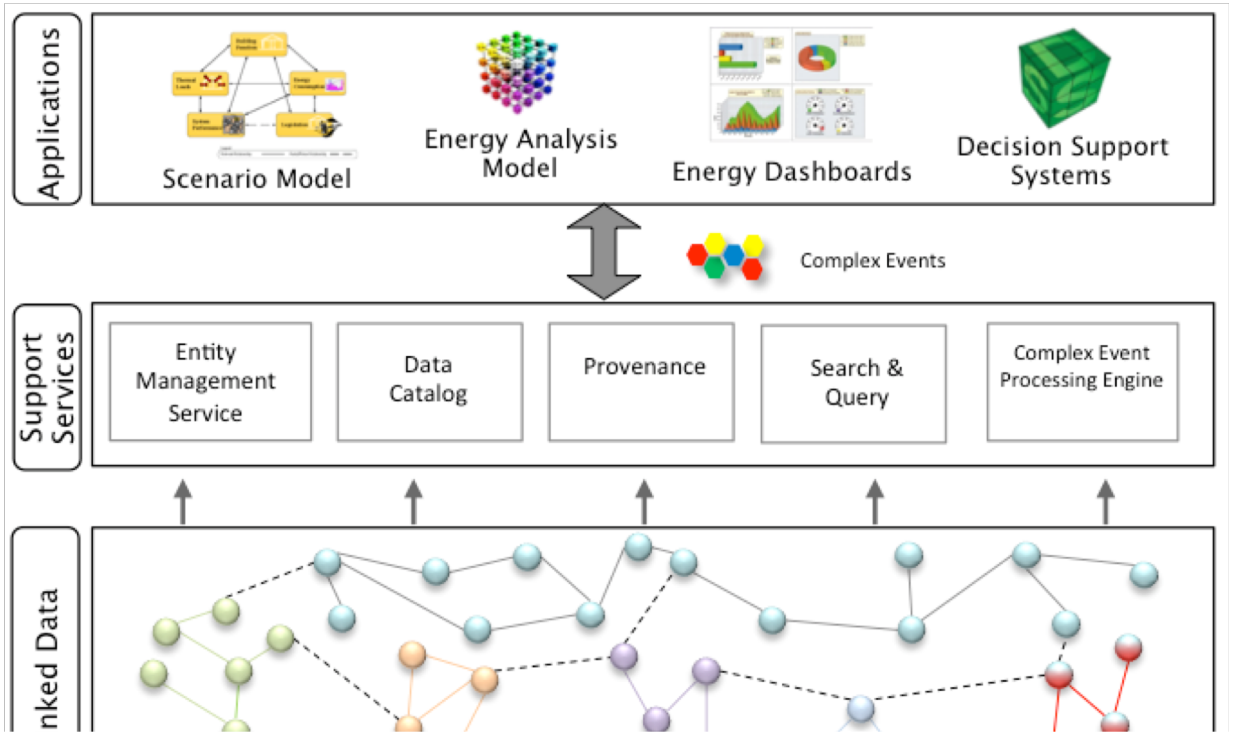
Linked data for buildings

Cross silo data exchange is the biggest challenge faced by data-focused companies today



O'Donnell, James, Edward Corry, Souleiman Hasan, Marcus Keane, and Edward Curry. 'Building Performance Optimization Using Cross-Domain Scenario Modeling, Linked Data, and Complex Event Processing'. *Building and Environment* 62, no. April (2013): 102–11. https://doi.org/10.1016/j.buildenv.2013.01.019.

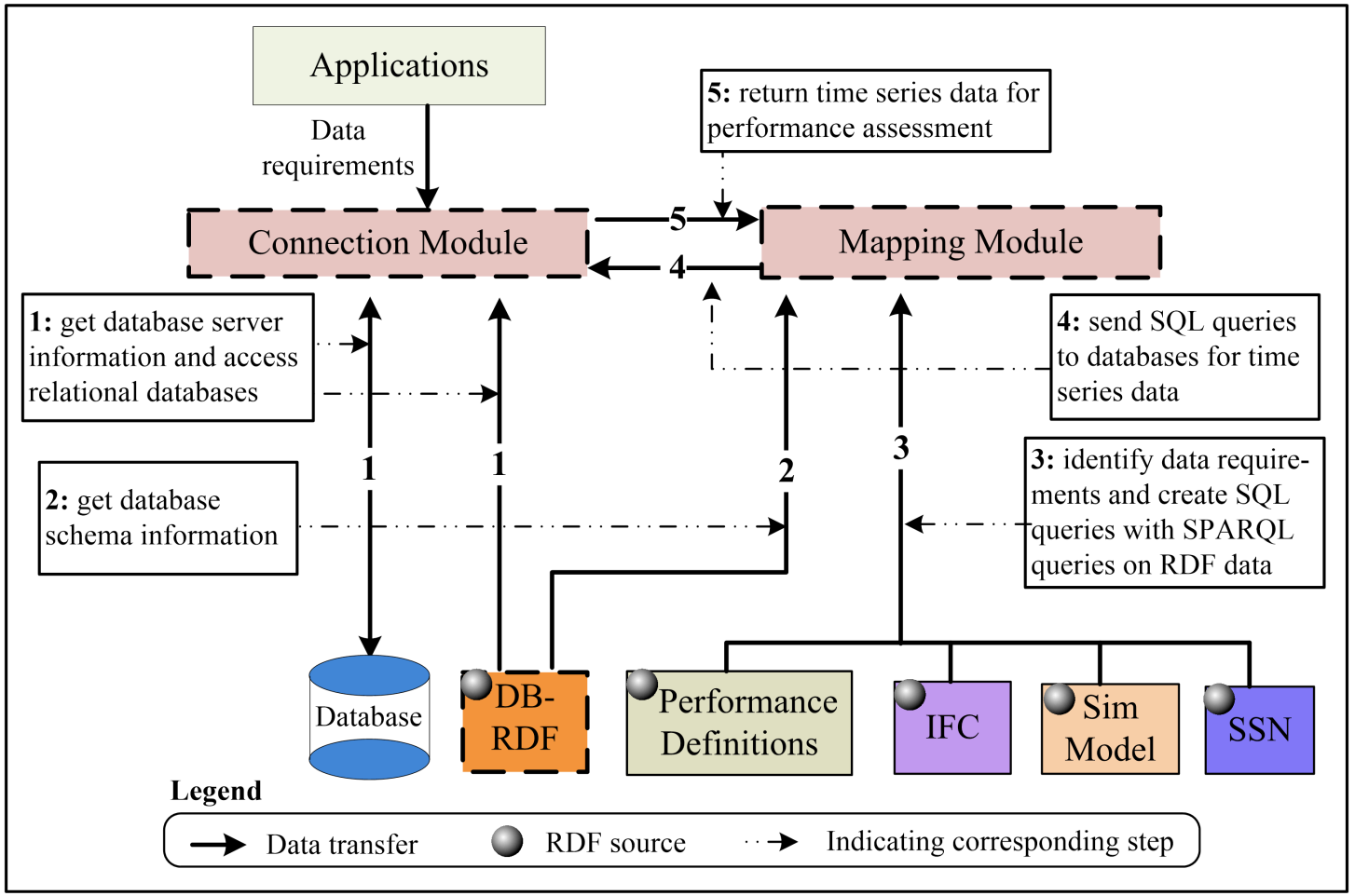
Cross silo information retrieval requires a comprehensive ICT infrastructure



Edward Corry, **James O'Donnell**, Edward Curry, Daniel Coakley & Marcus Keane, *Using cross-domain building and social media data sources to drive occupant satisfaction and optimise building performance*, submitted to Advanced Engineering Informatics, August 2013.

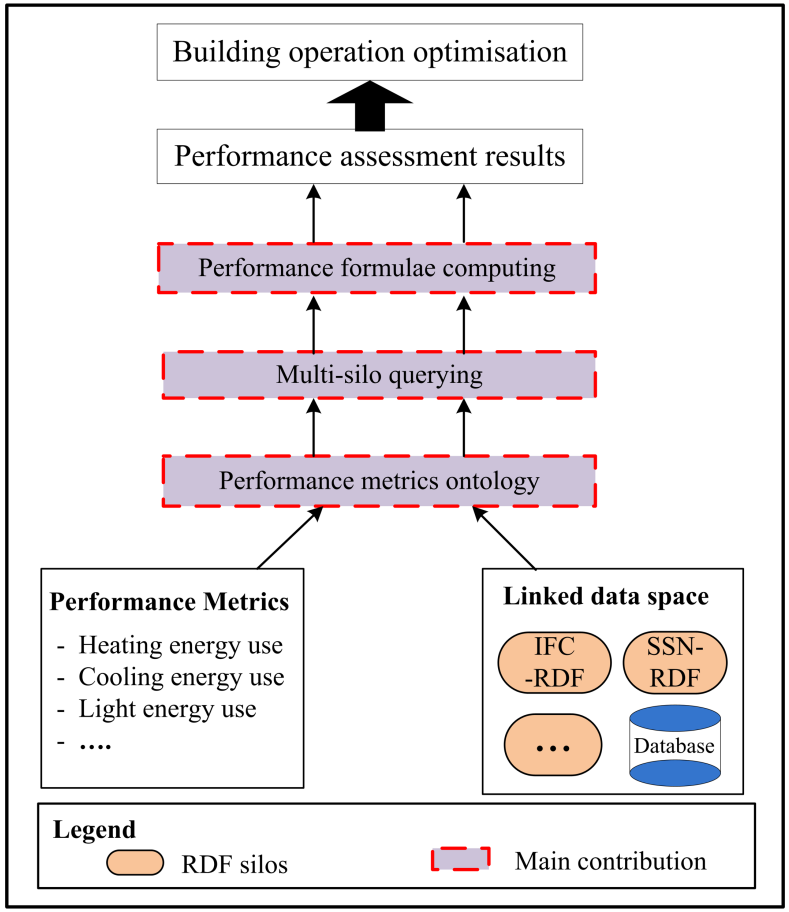
Edward Curry, **James O'Donnell**, Edward Corry & Souleman Hassan, *Using Multi-Domain Data to Optimise Building Operational Performance: A Linked Data Approach to Interoperability*, Advanced Engineering Informatics Vol 27, Issue 2, pp 206 – 219, 2013.

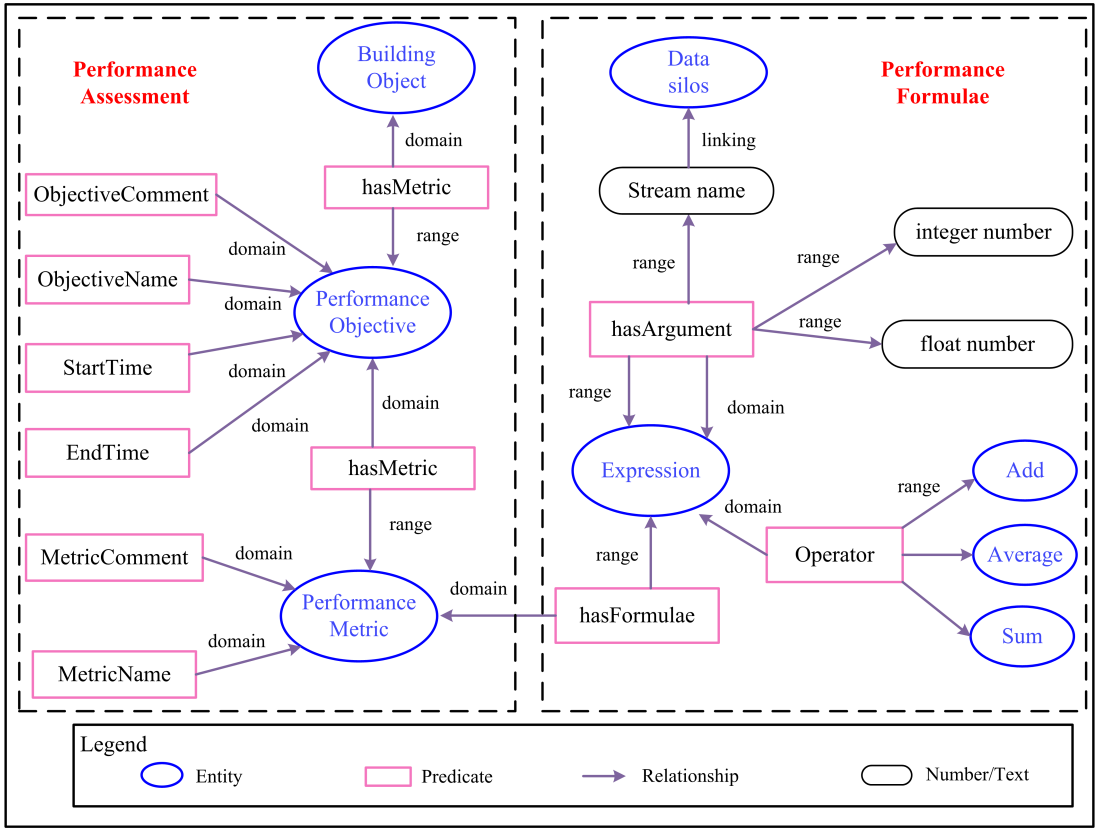
A hybrid-archetecture that integrates multiple data sources (RDF) with time-series databases (SQL)



Hu, Shushan, Edward Corry, Edward Curry, William J. N. Turner, and **James O'Donnell**. 'Building Performance Optimisation: A Hybrid Architecture for the Integration of Contextual Information and Time-Series Data'. *Automation in Construction* 70 (October 2016): 51–61. https://doi.org/10.1016/j.autcon.2016.05.018.

Integrating performance evaluation criteria and OpenMath (RDF) with time-series databases (SQL)

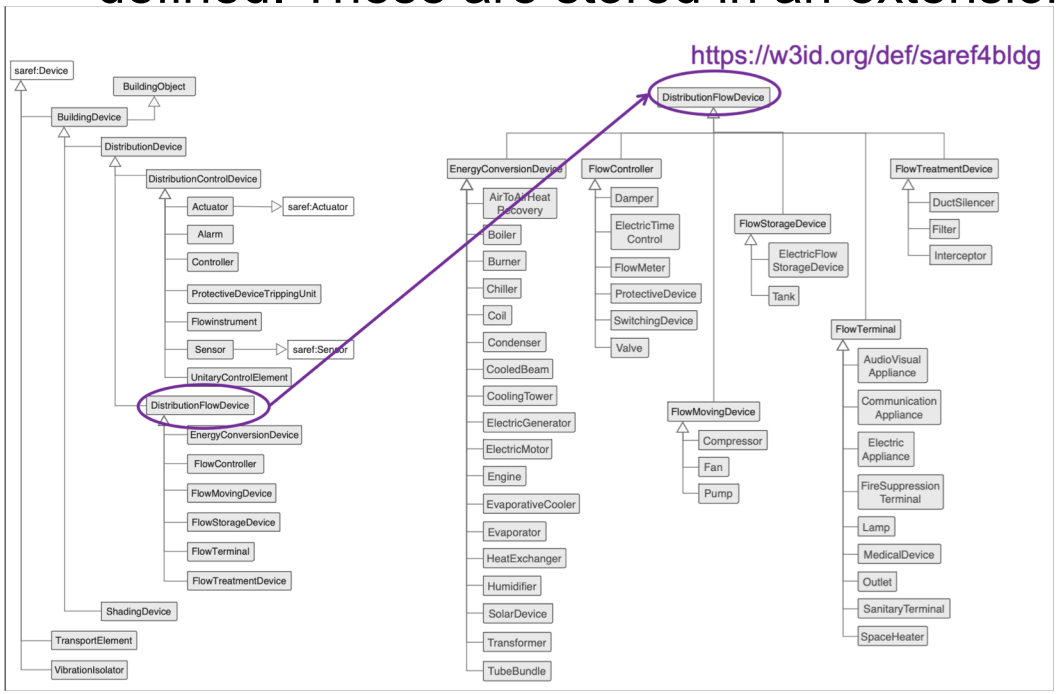




Hu, Shushan, Edward Corry, Matthew Horrigan, Cathal Hoare, Mathilde Dos Reis, and **James O'Donnell**. 'Building Performance Evaluation Using OpenMath and Linked Data'. *Energy and Buildings* 174 (1 September 2018): 484–94. https://doi.org/10.1016/j.enbuild.2018.07.007.

SAREF (use of linking rather than redefinition)

- SAREF links existing standards from different smart appliances based on different standards.
- It exposes a core model that links smart appliances at a high level
- Where different domains require more information, extensions are defined. These are stored in an extensions registry.



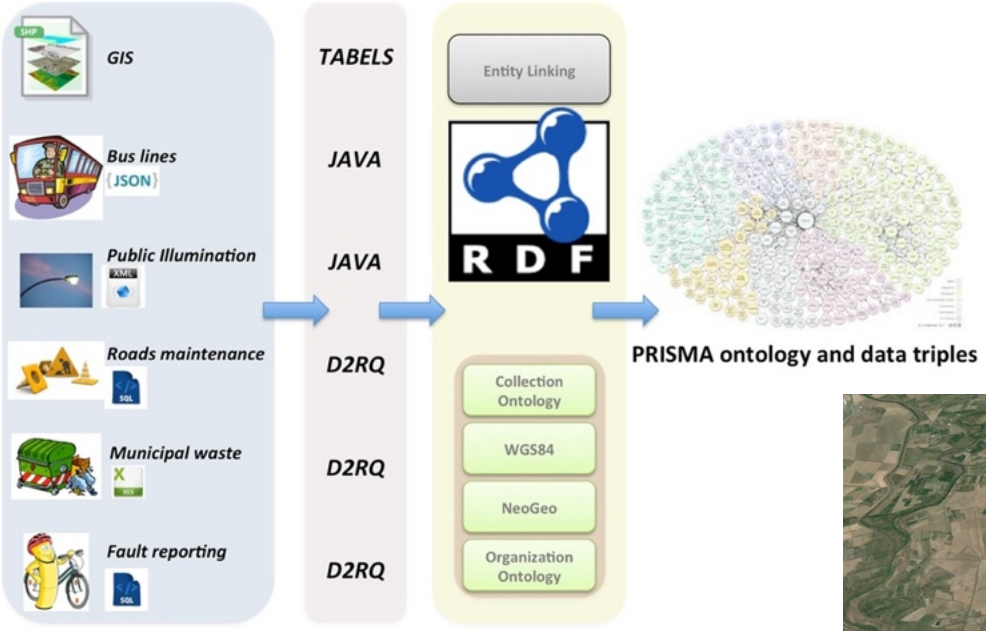
Result of Natural Language Processing across IFC classes to detect devices, along with use of Wordnet for class definitions to further detect devices.

Discovered entities built into a hierarchy of building devices.

Villalon et al, Extending the SAREF ontology for Building Device and Topology in: LDAC 2018, London

LINKED DATA FOR NEIGHBOURHOODS, DISTRICTS AND CITIES

Linked Data Generation in Smart Cities

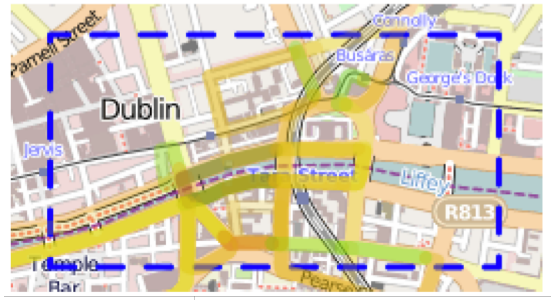


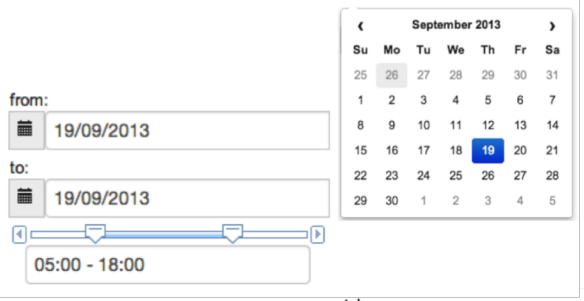
PRISMA ontology and data triples

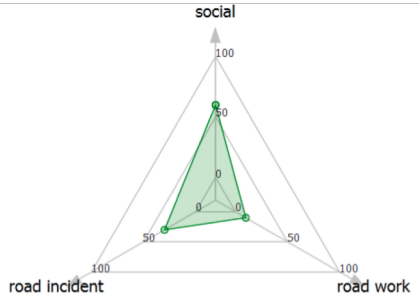
PRISMA Semantic Data Model

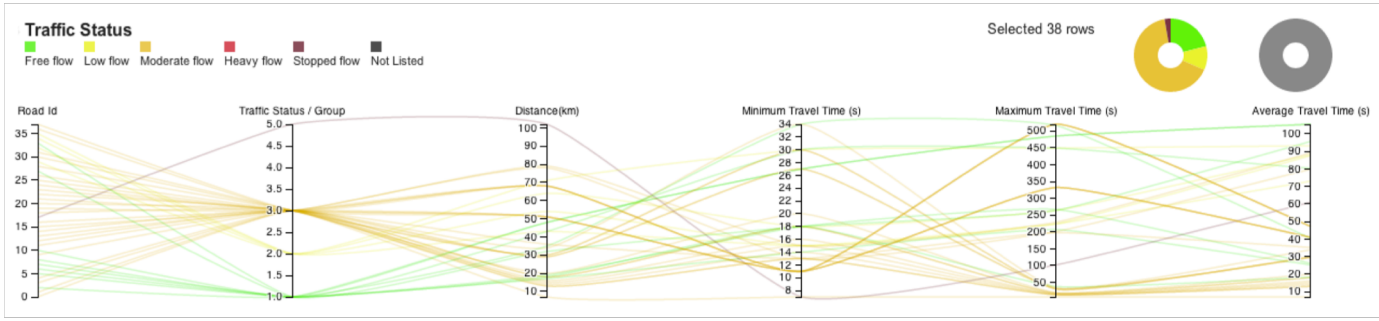
Consoli et al, Producing Linked Data for Smart Cities: The Case of Catania, In: Big Data Research, 2017

Semantic Traffic Analysis and Reasoning for CITY



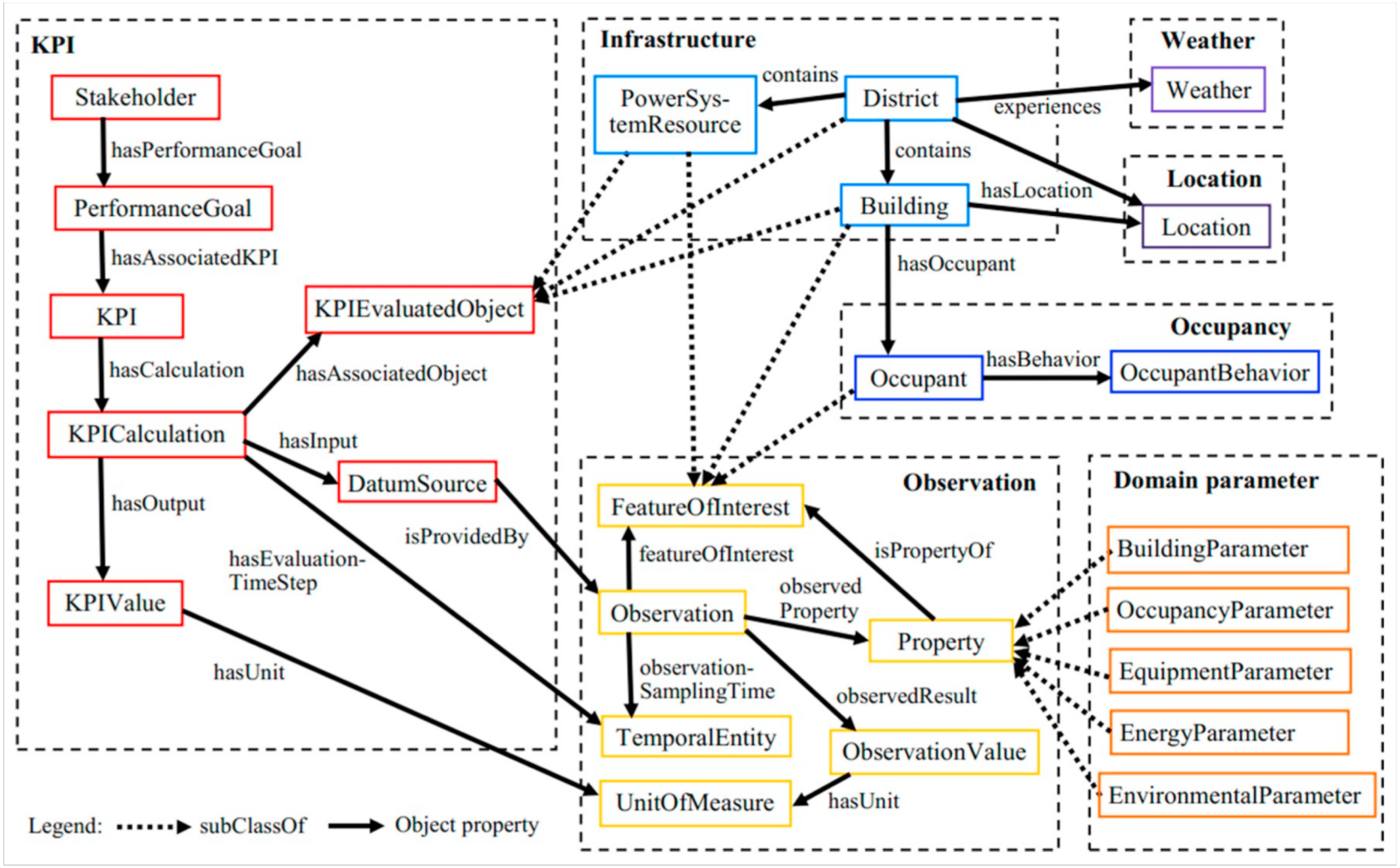






Lecue et al, STAR-CITY: semantic traffic analysis and reasoning for CITY, in: Proceedings of the 19th international conference on Intelligent User Interfaces

Overarching concept for an energy management Key Performance Indicators Ontology



Li, Yehong, Raúl García-Castro, Nandana Mihindukulasooriya, **James O'Donnell**, and Sergio Vega-Sánchez. 'Enhancing Energy Management at District and Building Levels via an EM-KPI Ontology'. *Automation in Construction* 99 (1 March 2019): 152–67. https://doi.org/10.1016/j.autcon.2018.12.010.

OPPORTUNITIES

Opportunities

- Use semantic web forms as opposed to PDFs….
- LDAC or Linked Data in Architecture and Construction 2019 is in Lisbon

EXTRA SLIDES

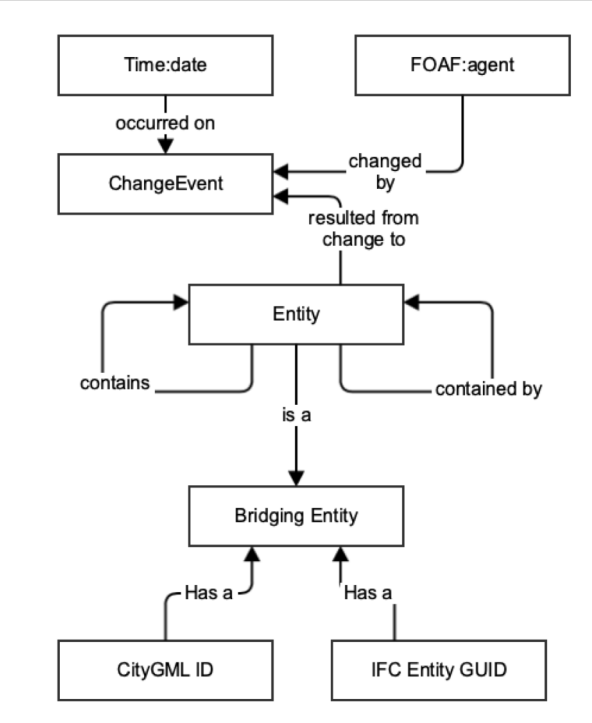
An Example

- Academic partner developed and managed a central data repository
  - Managed by a single software-competent personnel;
  - Developed in parallel with a series of software clients that created/used various stored data to run simulations in a variety of domains. Data was accessed through JSON APIs;
  - Clients development was in parallel – with various late-stage modifications required;
  - All client induced data changes flowed to server, and imposed collateral development on server-side with consequent knock on to clients;
- Linked data provides a means to implement a centralized meta-data repository that describes data and how to access it and a de-centralized set of data stores
  - Decoupled development and data means robust development process and flexible, reusable system

Issue 1 - Diverse formats and data exchange

- Problem – diverse formats have evolved to manage various aspects of construction related information. Usually, data must be extracted from these different standards requiring entity-matching. These range of standards used vary on a per-project basis. It is:
  - Expensive to create custom ontologies to manage project specific data exchange requirements
  - It is also expensive to transform data from a variety of standards and import it for use into a single standard
- Solution – create a simple (i.e. generic) common schema for the project that describes common points in different standards and then revert to data extraction from the original standards

Simple data translation schema



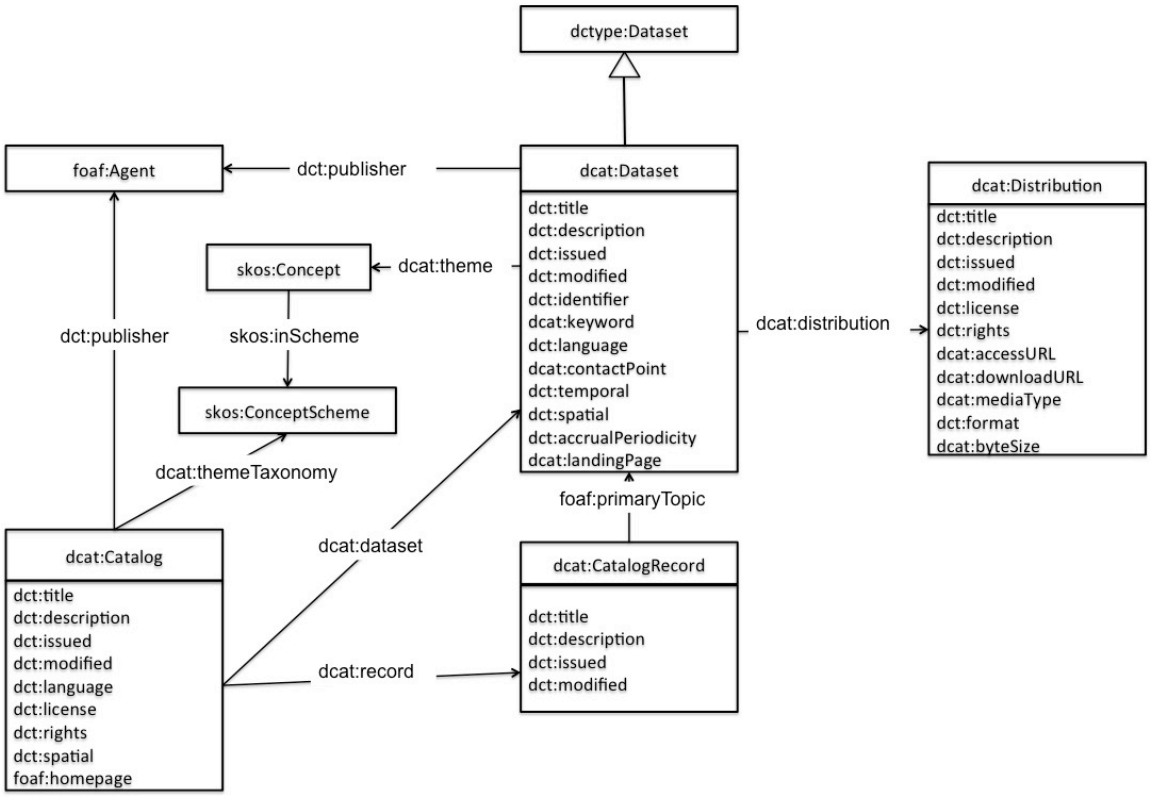
- Example of a simple translation schema.
  - Required combination of data from Urban (CityGML) and Building (IFC).
  - Created a hierarchy from urban level construct down to building level.
  - Where common entities existed, a bridging element was created, in this case, a building with a CityGML and IFC Entity GUID used as a translation point.
  - Now rules create queries that span multiple systems and extract required data into required format using translation schema as a guide.

Issue 2 – Data Access Control

- Problem – centralizing data can mean loss data control and associated code. This can cause commercial concerns as well as legal, placing constraints on potential of projects.

- Solution – Using DCAT as a centralized meta-data store about the datasets, project members can access information about information data is controlled on partners' own computers
  - Data remains 'findable' and can be reasoned about
  - Data access is controlled

Linked Data - DCAT



- Presents a semantic model of published datasets
- Acts as a catalogue and can be centralised for a project
- Distribution accessURL and downloadURL can be used to access the dataset and its RDF/Onthology metadata
- Actual data remains on the publisher's server and so owner maintains a control
- Numerous sources can be accessed and data mashed up. This does not have to be implemented on a central server, but rather at the point of use

Issue 3 – Cost of Implementation

- Problem - tightly coupled three-tier architecture:
  - Modifications are potentially expensive across the project
  - Especially expensive for the data store

- Solution – Using DCAT, centralize meta-data and distribute data to the data producer where data can be organized under the hood to suite their requirements while public interface can be used by clients to access data.
  - Data remains actionable for all participants
  - Data for any particular client can be easily repurposed in another project
  - Distributes cost of development to partners
  - Eases complexity in project management and implementation