Data Models and Semantic Interoperability

Moderator: Dave Raggett, W3C, Create-IoT, GATEKEEPER & Boost 4.0

email: dsr@w3.org

IoT Week, Aarhus, 20th June 2019

Five Talks – 15 minutes each

- *Introduction and Scene Setting*
  - Dave Raggett, W3C
- *AIOTI WG3 work on Semantic Interoperability*
  - Martin Bauer, NEC Labs Europe
- *ETSI STF 547?*
  - Mahdi Ben Alaya, Sensinov
- *Semantic Recipes and Schema.org*
  - Darko Anicic, Siemens
- *NGSI and RESTful API for IoT context management*
  - José Manuel Cantera Fonseca, FIWARE
- *Questions and Answers*

Introduction

- The big opportunities for the IoT are for services that combine data from multiple sensors and information sources, and across multiple ecosystems
  - Smart manufacturing as an example
  - Relationship to Digital Transformation in general
- Open markets of services
  - Enabling suppliers and consumers to find each other, and to reach agreement, e.g. on terms & conditions, payments, i.e. smart contracts
  - What’s needed to support security and trust
- This involves the exchange of data and metadata
- This session will focus on what’s needed to share data models and establish interoperability

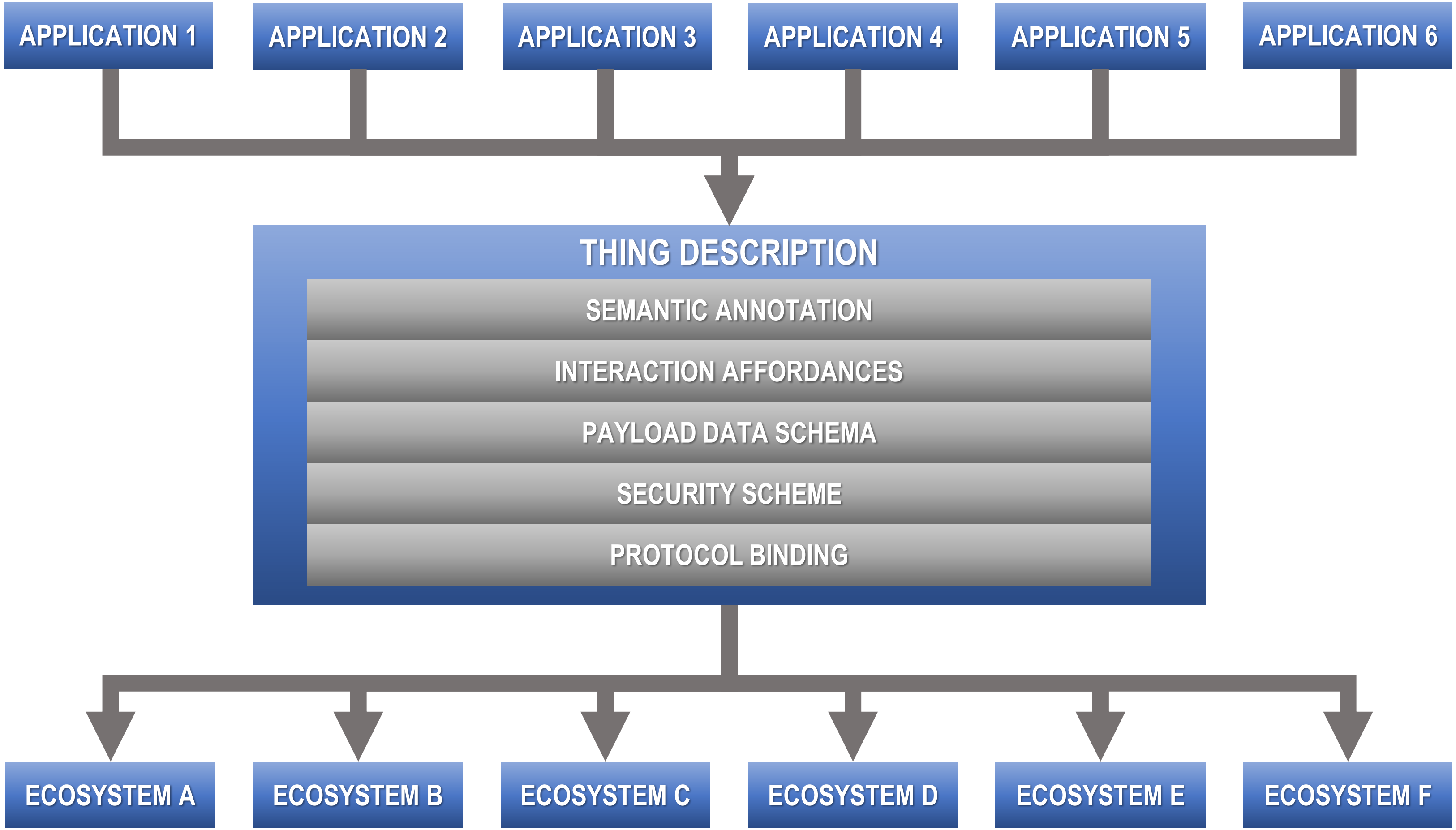
The Perils of Fragmentation

- Many different IoT technologies
- Proprietary ecosystems controlled by single vendors
- Confusing variety of standards and standards development organisations
- This results in silos that are expensive to bridge
- Moreover, the resulting complexity makes for brittle solutions
- **The IoT cannot realise its true potential until we solve this**

Web of Things as the solution

- Decouples services from underlying IoT technologies and protocols
  - Makes it easy for applications to integrate data from different ecosystems
- Things as digital twins for sensors, actuators and information services
  - Live objects, rich descriptions and reasoning/simulations
- Each thing is identified by a URI* that dereferences to metadata that describes the kinds of things and how they are exposed to client applications as local software objects with properties, actions and events, independent of the location of the sensors and actuators
- W3C's Web of Things Working Group is developing a standard around JSON-LD and expects to have this done by mid-2019
  - Currently at W3C Candidate Recommendation status
    - https://www.w3.org/TR/wot-thing-description/

* A *URI* is a Web address, e.g. *https://example.com/p23*

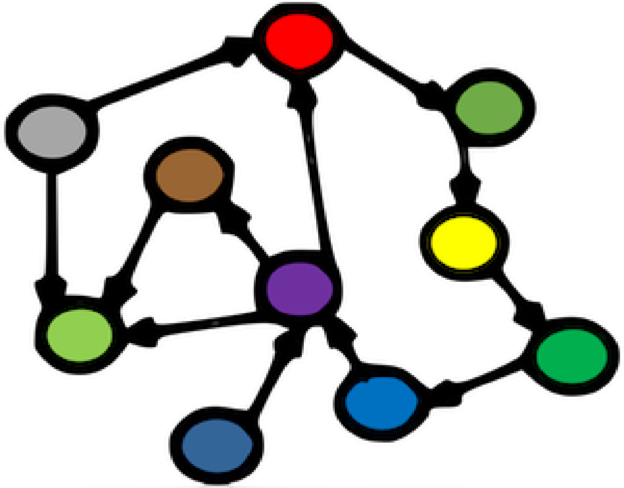

APPLICATION 1
APPLICATION 2
APPLICATION 3
APPLICATION 4
APPLICATION 5
APPLICATION 6
THING DESCRIPTION
SEMANTIC ANNOTATION
INTERACTION AFFORDANCES
PAYLOAD DATA SCHEMA
SECURITY SCHEME
PROTOCOL BINDING
ECOSYSTEM A
ECOSYSTEM B
ECOSYSTEM C
ECOSYSTEM D
ECOSYSTEM E
ECOSYSTEM F

How do I know what you mean?

- Interoperability requires agreement at multiple levels, including protocols, data formats, data models and the meaning of data
- Example: a sensor reading that is a floating point number that denotes a temperature value in degrees Kelvin. We also need to know what this measurement applies to, e.g. “reactor chamber 7”.
- We can state this information as metadata using agreed vocabularies of terms – also known as “ontologies”
- Different ontologies may be at widely varying levels of maturity and designed for different needs
- This comes with the challenge for mapping data between vocabularies used by different communities, loosely speaking, like translating between French and German

Graph Data and Knowledge Graphs

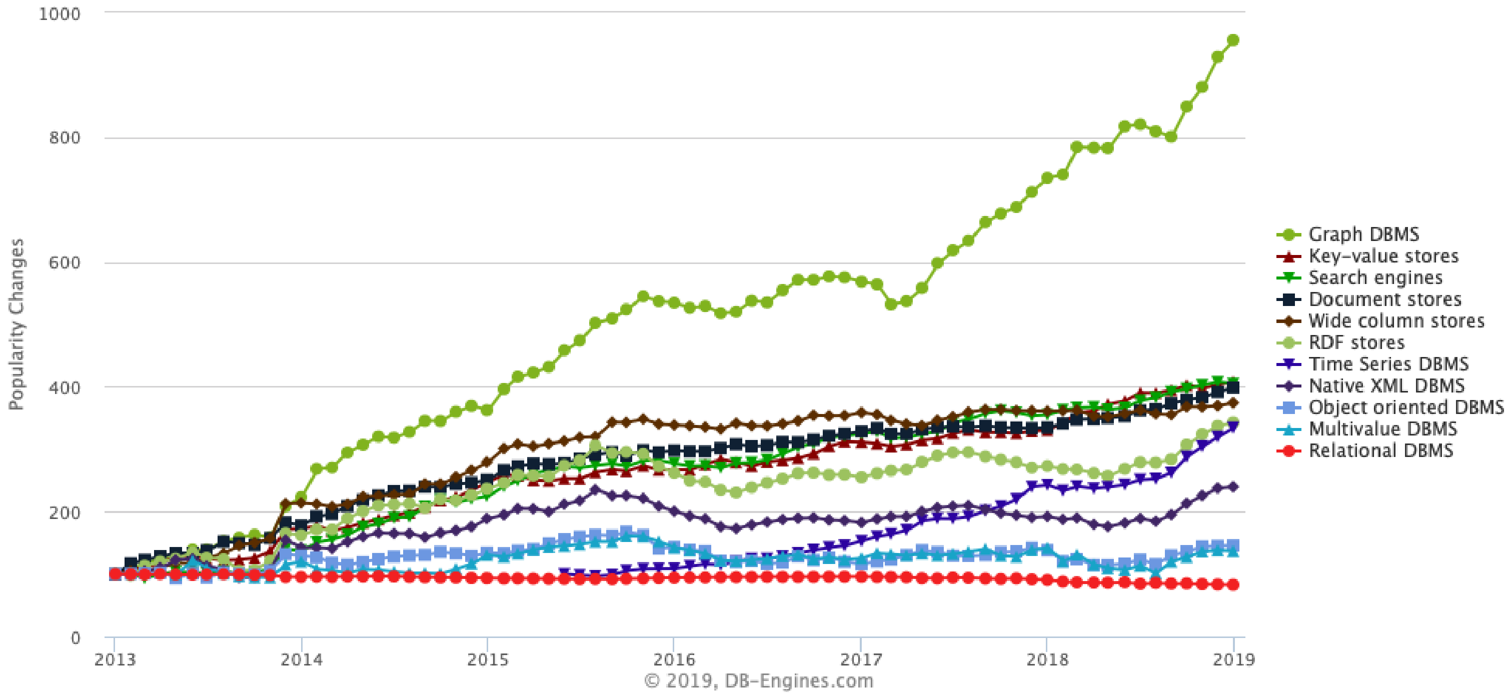
- Graphs are a very flexible way of representing relationships
- Applicable to both data and metadata
- Can be reasoned over with rules
- Can be processed with efficient graph algorithms
- Knowledge Graph is a term for graphs that describe concepts and relationships, often in the form of taxonomies
  - Popularised by Google for smart search results
  - Very relevant to businesses for enterprise wide data management



From tabular data to graph data

- Different traditions for data
  - Tabular databases: SQL/RDBMS
  - Graph databases: nodes and edges
  - PDF, CSV and spreadsheets
- We're in the midst of a shift to graph data
  - Faster than using SQL and associated JOIN operations
  - Better suited to integrating data from heterogeneous sources
  - Better suited to situations where the data model is evolving
- But lack of interoperability across commercial graph databases
- Emerging role for knowledge graphs as the basis for integration and governance across the enterprise
  - Addressing the lack of metadata for many data sources

Rapid Growth in Graph Databases



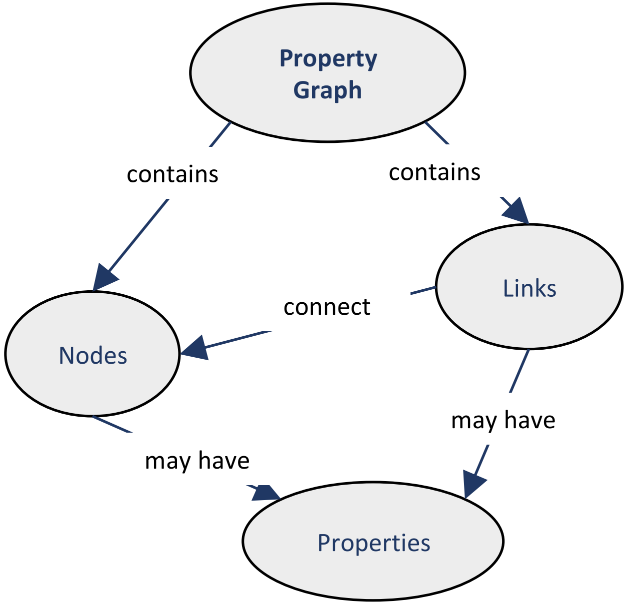
Courtesy of dbengines.com

RDF as the atomic foundation for data

- W3C's framework for Graph Data
  - Binary named directed relationships between concepts
    - <subject, predicate, object>, e.g. Mary loves Jack
- Global identifiers for concepts and predicates
  - HTTP URLs and URNs e.g. urn:tdm:aws:property:switch
- HTTP based identifiers can be dereferenced for more information
  - HTTP content negotiation to select between formats
    - e.g. HTML, RDF/XML, Turtle, N3, JSON-LD, …
- Mature suite of standards, e.g. RDF core, OWL, SPARQL, Turtle, JSON-LD, SHACL, DCAT
  - Lots of work on ontologies at W3C and elsewhere

Property Graphs*

- Framework for representing data and metadata with a graph of nodes and links
  - Both nodes and links may have name/value pairs
    - Otherwise referred to as “properties”
- Link annotations are very useful
  - Temporal, e.g. the time interval, link is effective
  - Spatial, e.g. location of sensor for a given reading
  - Provenance, e.g. the identity of the sensor
  - Quality, e.g. error bars on sensor reading



** Others know this as objects, frames or chunks; also related to the web of things*

With thanks to graphdatamodeling.com

W3C Graph Data Workshop – March 2019

https://www.w3.org/Data/events/data-ws-2019/

- Creating bridges across RDF, Property Graph and SQL communities
- Understanding requirements for interchange across graph databases
- Seeking alignment of graph query languages
  - SPARQL, GraphQL, Gremlin, AQL, Cypher, PGQL, SQL PGQ, …
- Discussion of extensions for RDF, SPARQL, JSON-LD, Turtle, N3
- Expected to lead to new standardisation work at W3C
- We're now launching a Business Group on Graph Standardisation
  - Collect use cases and requirements across domains
  - Guide technical standardisation work to match business needs
  - Liaisons with other SDOs and industry alliances

Easier RDF Initiative

- Value of RDF for graph data has been well proven, in many applications, over the 20+ years since it was first created
- But RDF is perceived as difficult to use, limiting its adoption
- We're seeking to make it easier for a much wider audience
- Guiding principles
    - Goal to make RDF, or some RDF-based successor, easy enough for *average developers*, who are new to RDF, to be consistently successful
    - Solutions may involve anything in the RDF ecosystem: standards, tools, guidance, etc. All options are on the table
    - Backward compatibility is highly desirable, but *less* important than ease of use
- See: https://github.com/w3c/EasierRDF

Generalising RDF – some ideas

- Tweaks to the core specifications after two decades of experience
  - Relaxing what can be used for subjects, predicates and objects
  - Base direction for string literals
  - Human language as just another property
- Allowing links to and from a single link or a set of links
  - Embracing Property Graphs without the need for reification
- Published identifiers versus nodes reachable via queries
  - URLs/URNs vs blank nodes
  - Named graphs vs unnamed collections of links
- Higher level framework designed around Property Graphs
  - Easier for most developers
  - Built upon the RDF core

Scaling up and up …

- How to make it easier to discover and learn about existing vocabularies?
- How to support work on vocabularies at widely varying levels of maturity?
- What is the sustainability model for developing vocabularies?
- Different communities will have different needs and different ways of talking about things
- This means that the IoT needs to support integration across such differences
- We will thus need scalable solutions for mapping data and services across communities and vocabularies
  - Peer to peer mappings vs mappings via an upper ontology
  - Context sensitive mappings
  - Potential role of statistical models
  - Glue code to span gaps
- Challenges for different approaches to identifiers
  - URLs, URNs, local names, paths from known names, etc.

Looking Further Out

A new paradigm for AI – bringing a human touch

- The Semantic Web focuses on deductive logic
- This doesn't work with incomplete, uncertain and inconsistent data which is likely to include errors
  - In short, it doesn't work well with raw real world data
  - Work around involving cleaning data using rough rules of thumb
    - The dark underbelly of Data Science …
- We will need to make use of ***rational* forms of reasoning** exploiting prior knowledge and past experience
  - Many forms of reasoning: deductive, inductive, abductive, causal, counterfactual, temporal, spatial, social, emotional, humour, empathy, compassion, …
- Hence the need to blend symbolic knowledge (graph data and rules) with computational statistics
  - Cognitive Psychology has shown how this can be realised
  - This will have a very profound effect on society
- This will allow us to create the ***Sentient Web***
  - Ecosystems of services with **awareness** based upon sensors, and **reasoning** based upon graph data and rules together with graph algorithms and machine learning

Synopsis

https://www.w3.org

- Huge opportunities for open markets of services that combine sensors, actuators and multiple sources of information
- Fragmentation + over emphasis on IoT edge technologies is holding this back
- W3C's Web of Things as solution, further work needed on open marketplaces
- The role of graph data and the need for agreement on meaning
- Making semantic technologies easier for the average developer
- What's needed for scaling up + best practices for vocabulary development
- Symbolic + statistical approaches to address complications of the real world
- Sentient Web = awareness + reasoning = digital twins + graphs + AI/ML
- W3C as the global forum for Web technology standards
- Please get involved as standards are critical for commercial exploitation
- For more information, please contact Dave Raggett <dsr@w3.org>