www.csiro.au

Semantically-enabled Standard Development

Laurent Lefort, CSIRO ICT Centre, Information Engineering Laboratory

Metadata Australia Conference, 26-27 May 2010

W3C Australia

CSIRO

Outline: W3C Semantic Sensor Network XG as a example of **semantically-enabled standard development**

- Motivations for the W3C Semantic Sensor Network XG
  - How do semantically-enabled standards fit in the current sensor network and sensor web standard landscape?
  - Expected benefits, challenges
- What's done in the XG
  - Primary goal of the XG: develop ontologies to describe sensors
    - To develop new types of sensor web and sensor networks applications: semantic mashups, provenance
    - To program sensor networks exploiting better knowledge about the sensors or the sensor network e.g. energy constraints
  - Secondary goal of the XG: develop semantic markup techniques
    - Define how ontologies can be used as a **complement** to existing XML-based standards (sensor web and sensor networks)
- Generalisation to other contexts
  - Commonalities and differences between sensing web and eGov

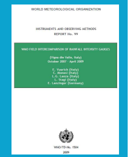
Why do we develop sensor ontologies? Sensor variability

- Vaisala Weather Transmitter WXT510/520
  - Uses an impact disdrometer
  - Rainfall, rainfall duration, rain & hail intensity
  - … Wind speed & direction, barometric pressure,
  - … Air temperature, relative humidity

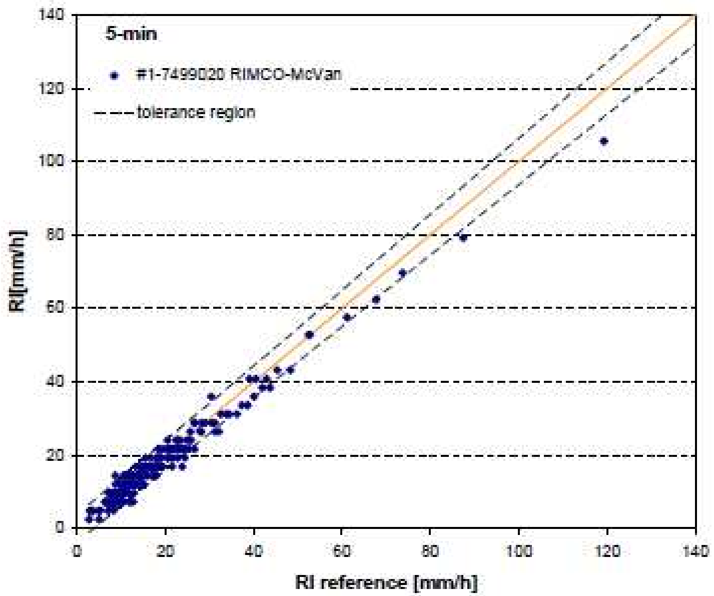
- RIMCO 7499 series
  - Tipping-bucket rain gauges (TBRG)
  - Rainfall

- Same function, one **difference**: **underestimation** threshold for high rainfall
  - Vaisala: okay until 80[1] to 200[2] mm/hr
  - RIMCO: okay until 300[1] to 500[2] mm/hr
    - References: [1] WMO study (see next page), [2] Manufacturer sheet

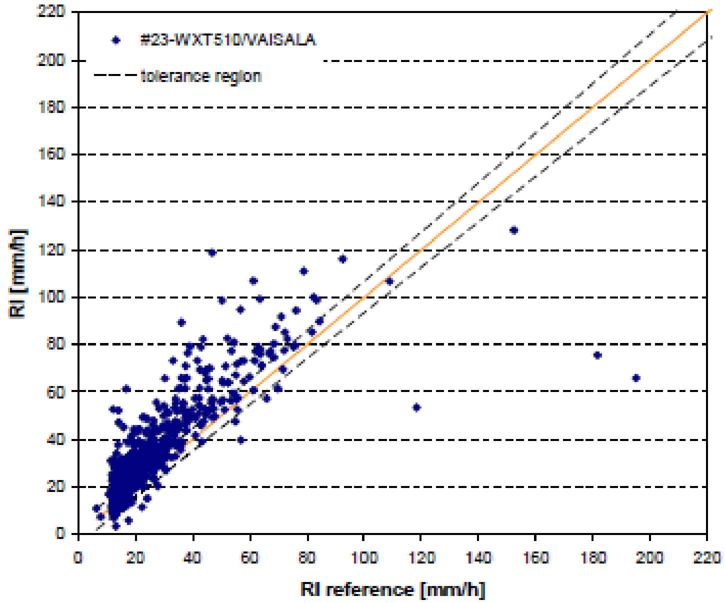
Reference: World Meteorological Organisation (2009) Field Intercomparison of Rainfall Intensity (RI) Gauges



- RIMCO



- VAISALA WXT 510



		Type	1-min RI resolution found (declared) [mm/h]	Measuring range found (declared) [mm/h]	Short comment	Performance in laboratory (constant flow)	Performance of 1 minute RI measurements
1	RIM 7499020 - Mc Van	TBRG with siphon	12 (12) insufficient	0 – 300 (0 – 500)	Under-estimation for high RI	***	***
23	WXT510-VAISALA	Impact disdrometer	0.1 (0.1) very good	0 – 80 (0 – 200)	Dispersion, over-estimation and under-estimation		**

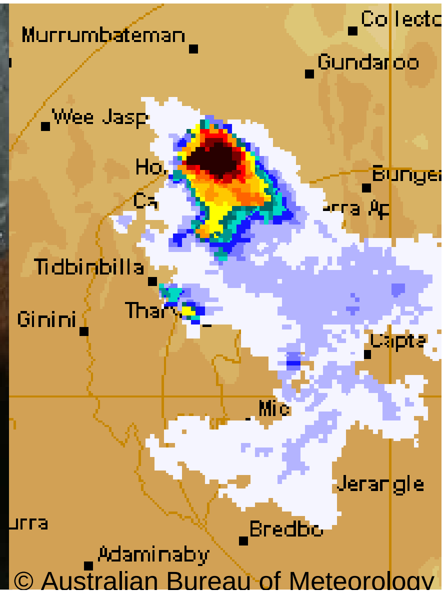
CSIRO

Why do we want to add semantic markup to data? Do you remember the hailstorm - 27 February 2007?

- Summary of Significant Severe Thunderstorm Events in NSW - 2006/07 (bom.gov.au)
  - "A **super-cell thunderstorm** at 10.30pm caused heavy rain and hail over Canberra City and Belconnen. Widespread hail up to 3cm size covered ground to **a depth of 20cm** causing 1 metre high hail drifts in Civic Centre. Up to **63mm*** was recorded at the Canberra Botanic Gardens in about 1 hour."
    - * note: the daily rainfall recorded in the BOM archives for the 28th February is 62.8 mm (bom.gov.au)


© Paul Hagon


© Paul Hagon



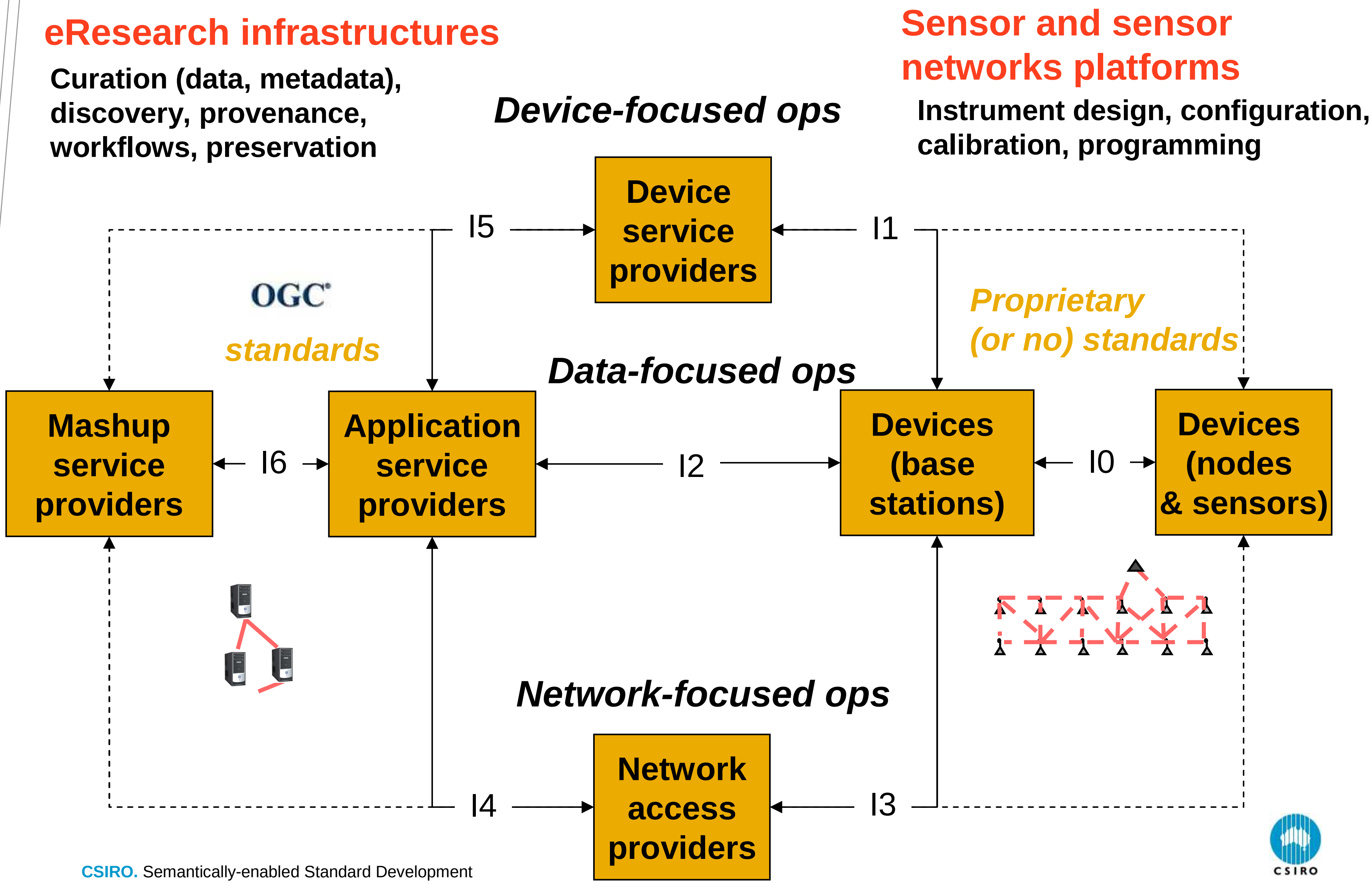
© Australian Bureau of Meteorology

Adding semantic annotations to find and use data

- Available information to estimate the severity of the storm (bom.gov.au)
  - Data: Daily record of rainfall for Canberra Botanic Gardens station
  - Text: Severe weather event description
    - Estimate of the length of the hail storm (related to the Botanic Garden observation)
    - Indication of the depth of the hail layer on the ground
  - Graphics: Radar image sequence

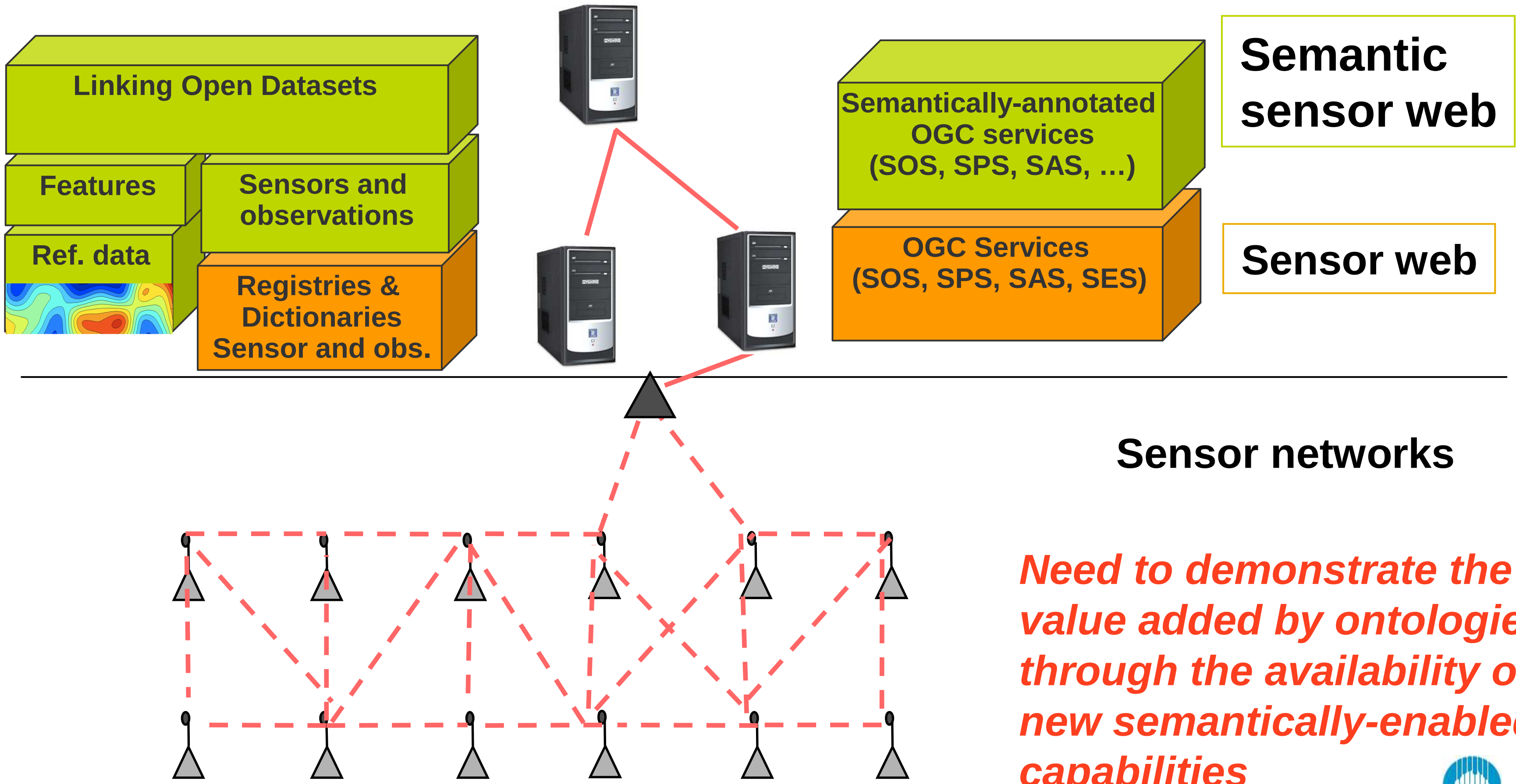
- Semantic annotations which could have helped
  - Information on the **type of storm** and on its trajectory over Canberra
    - Captions or metadata added to the radar image sequence
  - Complement to the weather observation
    - **Provenance** of the 1hour estimate for the storm duration
    - **Uncertainty** attached to the rainfall record
  - **Linkage** to other sources of information
    - How the rain gauge performs under such conditions (operational range of the class of rain gauges used by BOM)
    - What water-equivalent-of-hail formula is applicable to this type of hail?

Sensor network and sensor web standard landscape



Ontologies for better apps (discovery, provenance, composition)

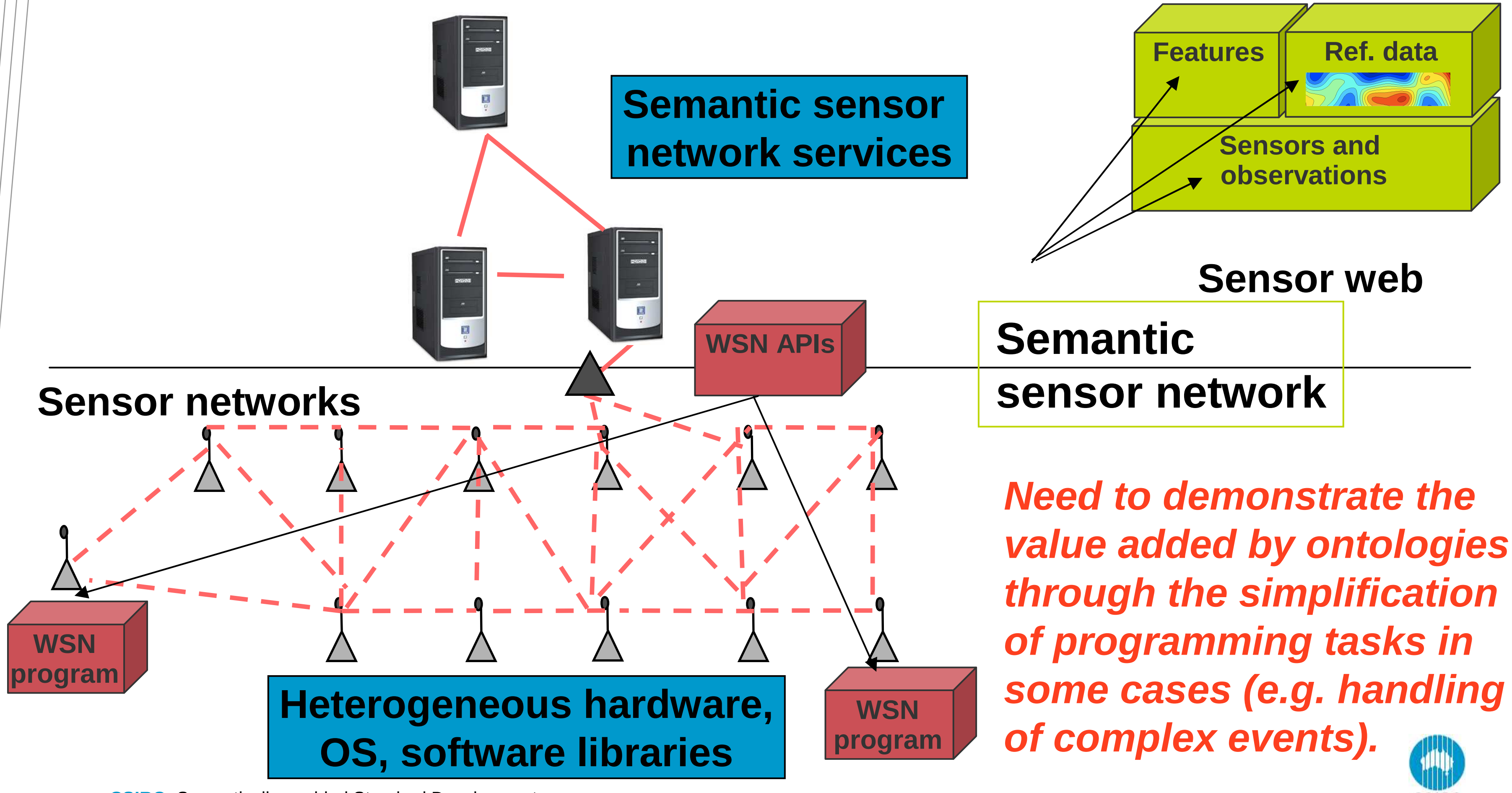
Motivation: new semantic discovery/composition/reduction capabilities (complete standards which are too generic e.g. OGC SWE services)



Need to demonstrate the value added by ontologies through the availability of new semantically-enabled capabilities

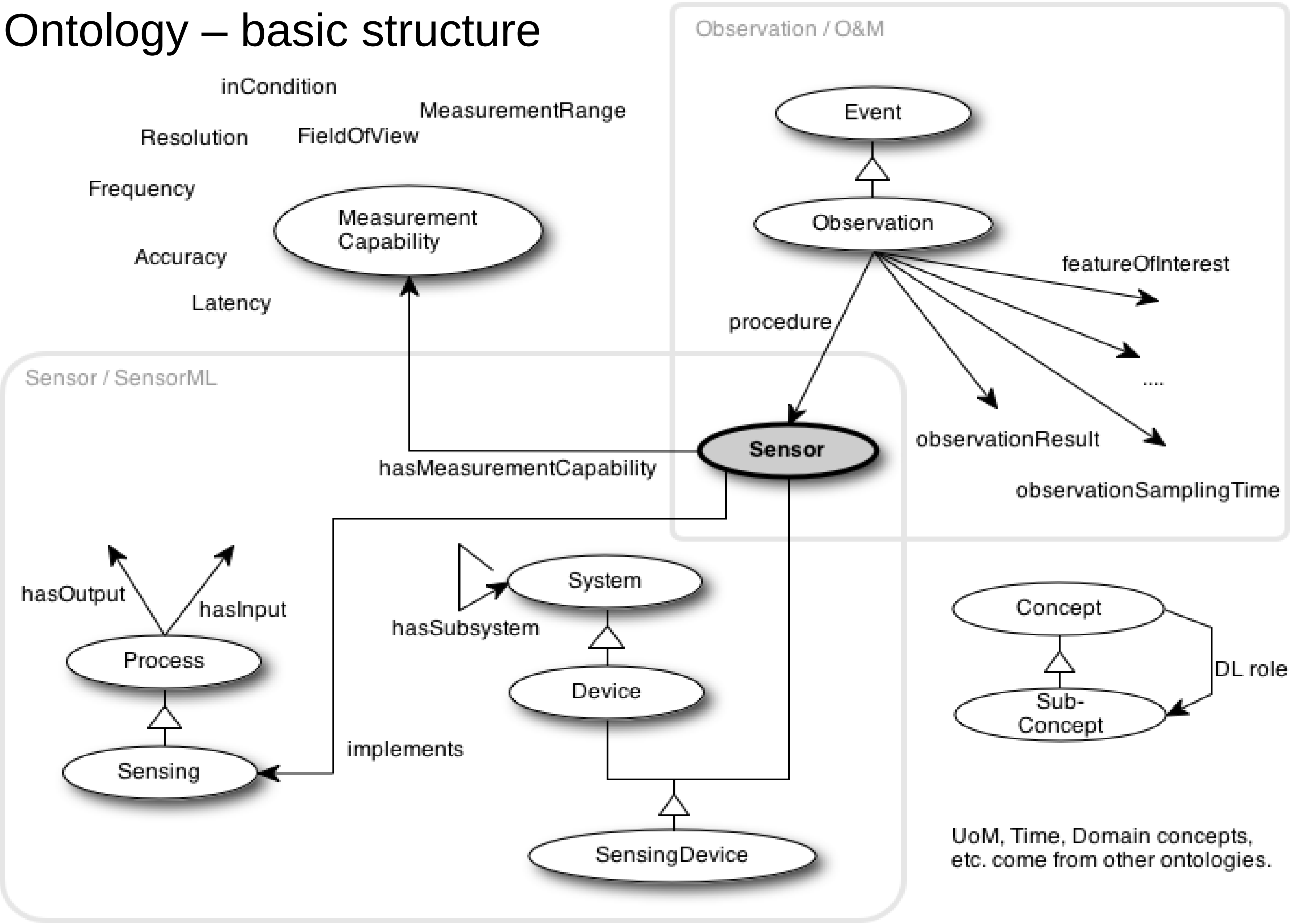
Ontologies for better sensor networks (macro-programming, diagnostic)

Motivation: develop new methods to generate programs for sensor networks which better exploit engineering analysis knowledge e.g. the operational range of the sensor, the energy constraints



Need to demonstrate the value added by ontologies through the simplification of programming tasks in some cases (e.g. handling of complex events).

Ontology – basic structure



Also adding:
physical properties, power use, connectors, lifetime, etc (of devices/systems)
mobility, availability, operational ranges, callibration, ...

Semantic annotation

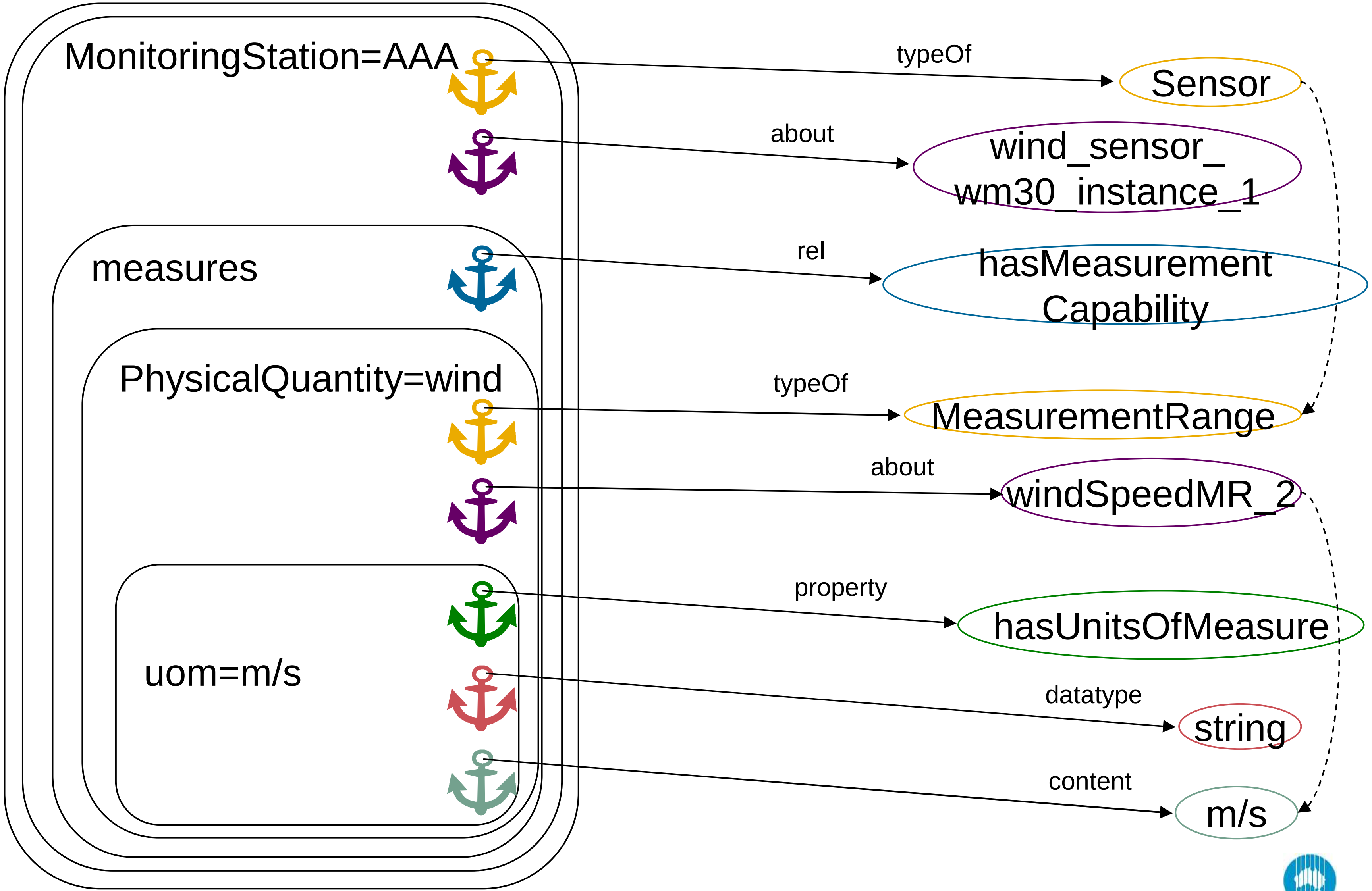
- **Semantic annotation** (for a document containing text or data): a web annotation which adds information to a web resource that is described in an ontology
- Can use **any** of the categories of definitions which can be included in an ontology

- Such annotations may be added to different types of content
  - XML, service descriptions file, HTML

Type of semantic annotations
Domain class (Sensor)
Domain instance (wind_sensor_wm30_instance_1)
Object property (hasMeasurementCapability)
Range class (PhysicalQuantity)
Range instance (windSpeedMR_2)
Datatype property (hasUnitsOfMeasure)
Datatype property type (xsd:string)
Range value (m/s)

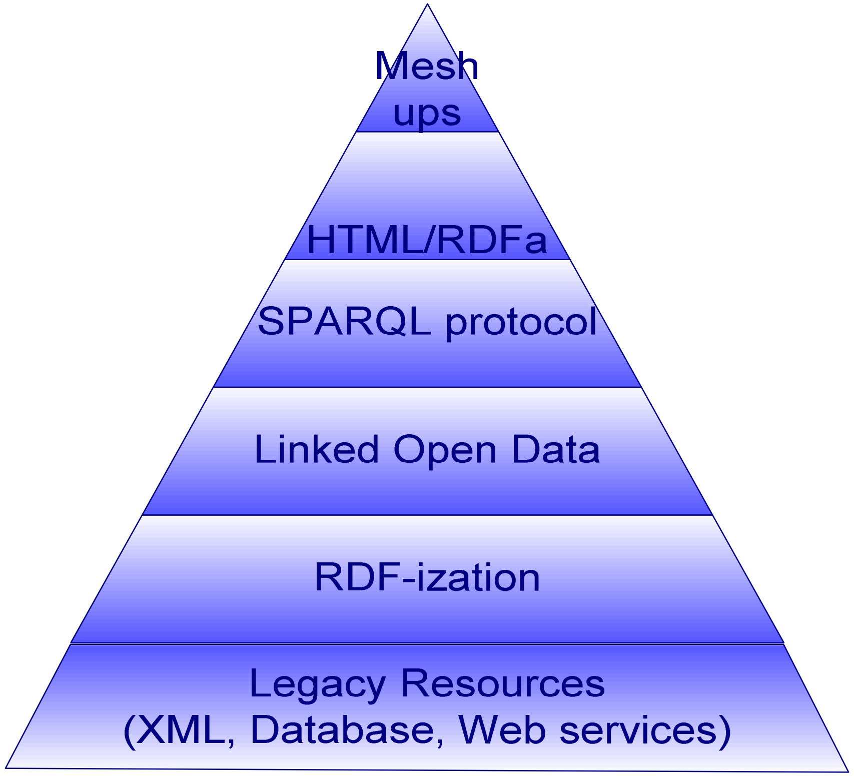
Example of semantic annotation

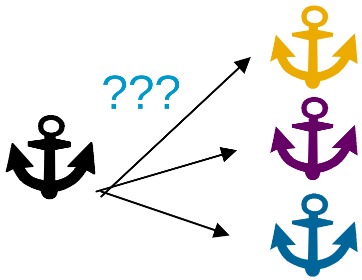
XML file | Ontology



What we have learned

- Priorities for ontology development:
  - Skeleton to add more links between different domain views
  - Placeholders for richer content (allow for future extensions)
- Multiple semantic markup solutions corresponding to multiple levels of the standard stack
  - RDFa is the most complete approach (benchmark)
- Issues related to the addition of semantic markup to existing XML-based standards
  - How to compare / combine solutions based on different standards families (XML, RDF and even HTML)
  - How to match the “classic use” of URNs (in OGC) to the **three types** of semantic web concepts (classes, properties and individuals)





Generalisation to other contexts

- Emerging trend (1): ontology-based standards (now)
  - RDF-only standards, using OWL or Linked Data (SPARQL)
    - W3C: Delivery Context ontology, Media Resource ontology, ontologies developed by Incubator Activities (XGs)
    - OGC GeoSPARQL (Geosemantics WG)
    - OASIS Quantities and Units of Measure Ontology Standard (QUOMOS)
- Emerging trend (2): **Semantically-enabled standards** (soon)
  - Mix of XML-based and RDF-based standards
  - Ontologies & semantic markup capabilities to enrich existing standards
- Generalisation to other contexts (eGov)
  - Rationale to adopt semantic web standards (Gov 2.0 TF project 5)
  - Commonalities and differences between sensing web and eGov
  - Where to use small (simple) or large (complex) ontologies?

Rationale to adopt Semantic Web standards
“What users are telling us” (source: gov2au project 5)

- User frustration, confusion of not knowing where to go
  - People feel **overwhelmed by the amount of data available** because they just can’t find what they want, not because the data doesn’t exist.
  - Information cannot be obtained without access to insider’s knowledge of how government agencies work and publish information
- Mismatch between offer and demand
  - Information published from the **agency-centric perspective**, rather than from an **holistic end-user perspective**.
  - Resources only aimed for **people with the technical skills** to actually get in and use it, **not for the average person** who has a real need.
- Agency frustration
  - **Enormous amount** of “legacy” data and information sitting within government departments in various formats which is of **enormous value** but is **not publicly accessible**, let alone known about.
  - Siloed approach to the provision of data and information within and between government departments - need **“some way to join the dots”**
- See Berkhout M., Rowland-Campbell, A. and Strahl, P. Gov 2.0 TF Project 5: Semantic Tagging of Government Websites Nov 2009

Commonalities and differences between Sensing Web and eGov (key requirements)

- **Same**: Give average users
  - the ability to find all the available data
  - the ability to use all the available data
- **Same**: Match information to end-user perspective
- **Same**: Provide some way to join the dots

- **Variable**: Give advanced users
  - the ability to find all the available data
  - the ability to use all the available data
- **Variable**: Handle large amounts of data

Commonalities and differences between Sensing Web and eGov

- **Same**: (too) many standards which are too generic
  - Develop ontologies which capture what is not enforced by the standards but which is useful
  - Extend the standard to allow (optional) semantic markup
- **Same**: risk of confusion because of the simultaneous use of multiple family of standards
  - Mashup, meshups and the “stack”
- **Same**: difficulty for decision-makers to understand the added value of SW technos (where to use small/simple or large/complex ontologies?)

- **Different**: in some cases (Health), the ontology (SNOMED) has been developed “independently” of the leading XML standard (HL7)

- **Variable**: existence of standards / solutions which provide alternative approaches to serve the same goal e.g. registry, catalogues, metadata profile
  - Three key criteria: linkability, completeness and automation

Large (complex) or small (simple) ontologies?

Semantic reduction →

Key requirements	Server-side	Client-side
Give average users the ability to find all the available data *(discovery)*	Large or small	Small
Give advanced users the ability to find all the available data *(discovery, provenance)*	Large	Large or small
Give average users the ability to use all the available data *(mashups)*	Large or small	Small
Give advanced users the ability to use all the available data *(composition, mashups)*	Large	Large or small
Match information to end-user perspective *(model-driven transformation)*	Large	Large or small
Handle large amounts of data *(macro-programming)*	Large or small	Small
Provide some way to join the dots *(linking)*	Large	Large or small

The difficult transition from Web 2.0 to Web 3.0

- What happens once gov[t] data resources are available?
  - Designers, developers and other experts building mashups (govhack, apps4nsw, app-my-state VIC) but use semantic web technos **very rarely**
- Report of the Gov 2.0 Taskforce
  - […] Information should be: free, easily discoverable, based on open standards and therefore machine-readable, properly documented and therefore understandable, licensed to permit free reuse and transformation by others
  - […] governments have a role to play in leading and encouraging the uptake of **Web 3.0 technologies** in support of greater innovation based on the reuse of public sector information and enhanced citizen/government interaction.
    - e.g. RDFa usage in data.gov.au
- But … risk of standard adoption for the **wrong reasons**
  - Tommie Usdin Standards considered harmful Balisage 2009 : *"**mindless** application of standards that are not applicable is harmful."*

Conclusions:
Adopt Semantic Web standards for the right reasons

- Semantic Web and XML standards are not mutually exclusive
  - Porting everything into RDF may not solve all your problems
- Standards Development Organisations should adopt common semantic markup / annotation methods to develop the next generation of XML-based standards
  - Requires better coordination between XML/Web services and Semantic Web activities @ W3C
    - Guidelines and also tools supporting hybrid standards

- Areas for further work
  - Tools for semantic mashups (meshups)
  - Multi-standard stack issues in W3C, especially the validation of data/documents using hybrid standards
  - Evaluation of strengths and weaknesses of existing XML standards “extension” methods (e.g. ISO 19115 Metadata profiles) against the proposed approach

CSIRO ICT Centre
Laurent Lefort @laurentlefort
Ontologist, Information Engineering Lab
W3C Australia Office @W3CAustralia

Photo: Web Directions

Semantically-enabled standard development

- ***Abstract**: Today, XML technologies are routinely used to define semi-normative versions of IT standards by a number of SDOs (Standard Development Organisations). Some SDOs (W3C, OGC or OASIS) have started to work on the addition of Semantic Web technologies (ontologies, RDF/linked data) to their standard development toolbox for a number of reasons: richer expressivity, web-ification of shared resources, basis for harmonisation.*
- *This presentation will cover the work done within the W3C Semantic Sensor Network Incubator Activity (SSN-XG) on the transition to semantically-enabled standards and discuss how machine-processable ontologies can be designed and used to upgrade services based on existing XML-based standards*

Semantic web technologies

- New capabilities for web-based applications
  - Ontology engineering: description logic
  - Linked data management: triple store, SPARQL
    - URI-based IDs and queries
  - Annotations / Metadata : RDF, RDFa
- Niche areas
  - Discovery, Provenance, Composition, Curation, Annotation
- Can help to address standard development gaps such as
  - Profiling of standards (which are “too generic”)
  - Curation of large scale vocabularies
  - Model-Driven development with Open World Assumption
  - Harmonisation of domain models

Replace existing standards

- Replace existing approaches
  - Controlled vocabularies in SKOS or OWL
  - Domain models migrated to OWL (instead of XML schemas or UML models)
- Pros
  - Technical factors: better management of vocabularies, URIs/APIs facilitating mashups
  - Social factors: ability to federate sub-communities (foundries)
- Cons:
  - Technical factors: SW technologies less effective for some tasks: e.g. closed models (including User Interfaces based on forms and data validation)
  - Social factors: SDOs reluctance to lose investment in non-SW models (UML) & schemas (XML)

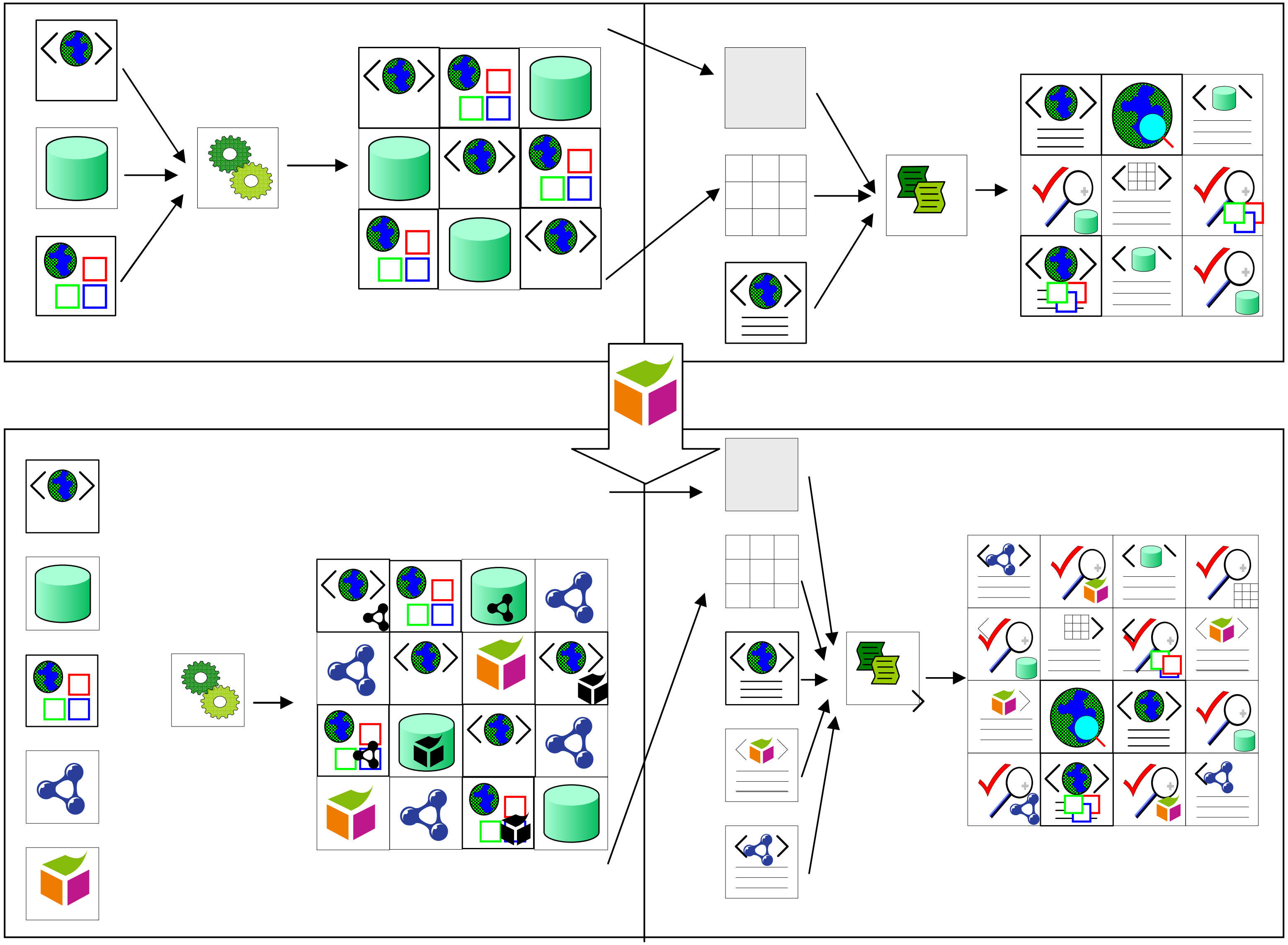
Upgrade (Semantically-enable) existing standards

- Upgrade existing approaches
  - Migration of vocabularies into SKOS or OWL-based resources hosted on triple stores
  - Semantic annotations to extend web services with lifting (XML-to-RDF) and lowering (RDF-to-XML) operations
- Pros
  - Technical factors: benefits of new capabilities, solutions to standard development gaps
  - Social factors: openness to all categories of contributors to standards
- Cons
  - Technical factors: complexity of standards stack relying on XML + RDF (+ HTML)
  - Social factors: lack of consensus on SW-enabled standard development

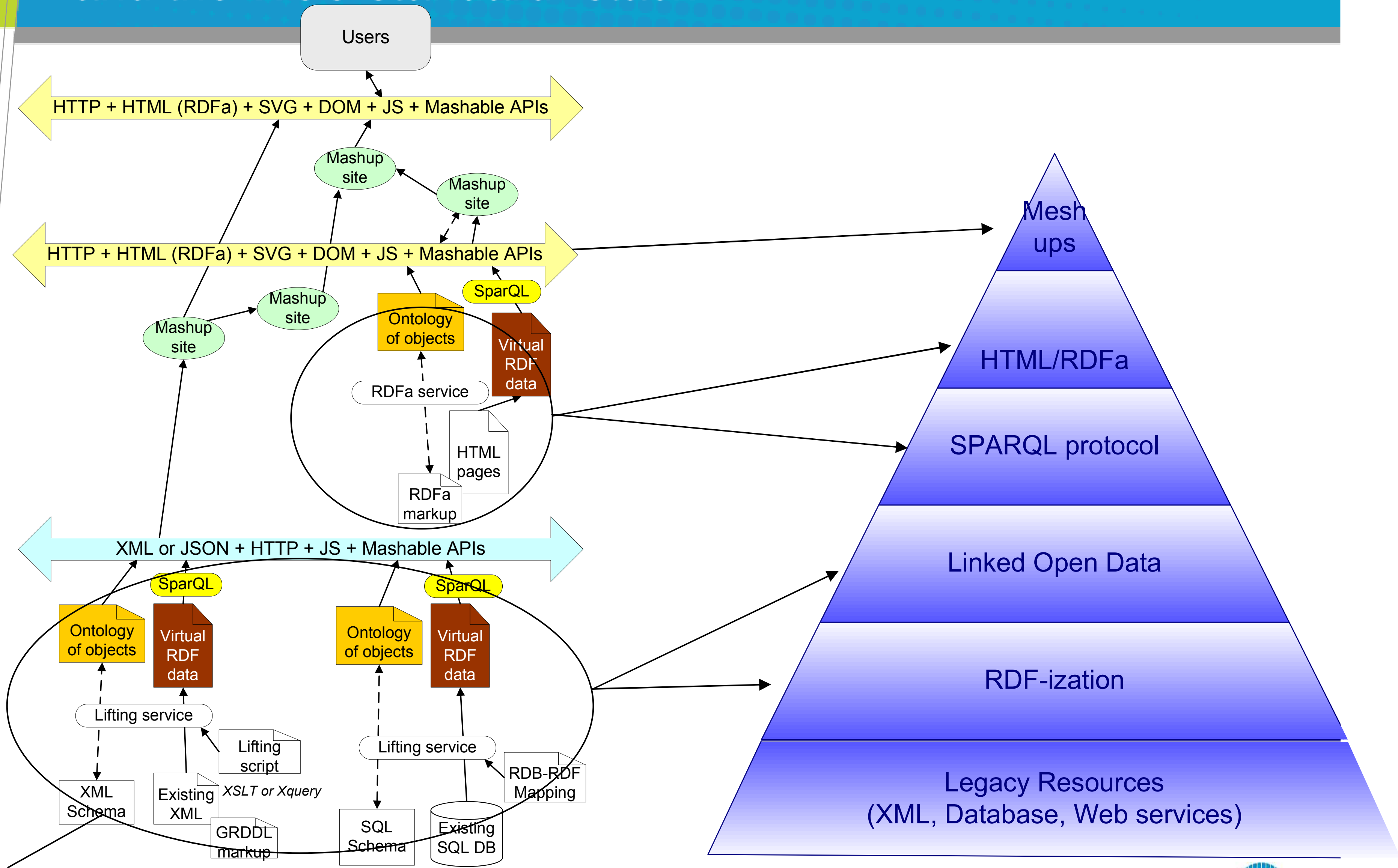
Do nothing

- Do nothing
  - Do not change existing standards
  - Develop new approaches useable separately (e.g. DBPedia)
- Pros
  - Technical factors: backward compatibility of future versions of standards
  - Social factors: keep it simple stupid and if it ain't broke, don't fix it principles
- Cons
  - Technical factors: harder maintenance of solution based on closed world assumption, difficulty to migrate to Linked Open Data cloud
  - Social factors: barrier for the development of multi-disciplinary standards

Semantic enablement: where to do it?



Semantically mashable semantic mashups and the W3C standard "stack"



Tim Berners-Lee, Cracks and Mortar W3C TPAC 2007

References

- WMO (2009): Field Intercomparison of Rainfall Intensity (RI) Gauges
- Bureau of Meteorology Summary of Significant Severe Thunderstorm Events in NSW - 2006/07 Bureau of Meteorology Observation of rainfall (bom.gov.au)
- Jones, K. F. Ice Storms in the St. Lawrence Valley Region, Technical Report Army Corps of Engineers Engineer Research and Development Center January 2003 (non peer-reviewed)
- Bureau of Meteorology Daily rainfall CANBERRA (AUSTRALIAN NATIONAL BOTANIC GA) 2007 data
- Bureau of Meteorology BoM Canberra Radar Loop - Rain Rate - IDR403 (via www.theweatherchaser.com)

- Semantic Sensor Network XG (ends in September 2010) http://www.w3.org/2005/Incubator/ssn/
- Lefort, L. Review of semantic enablement techniques used in geospatial and semantic standards for legacy and opportunistic mashups in Proc. of the 5th Australian Ontology Workshop (AOW 2009) paper and slides
  - Plus Review of Semantic Annotation techniques: March 2010 SSN XG slides

- RDFa http://www.w3.org/2010/02/rdfa/
- Usdin, T. Standards considered harmful Balisage 2009

- Berkhout M., Rowland-Campbell, A. and Strahl, P. Gov 2.0 TF Project 5: Semantic Tagging of Government Websites Nov 2009
- Government 2.0 Taskforce Engage: Getting on with Government 2.0 Report of the Government 2.0 Taskforce December 2009