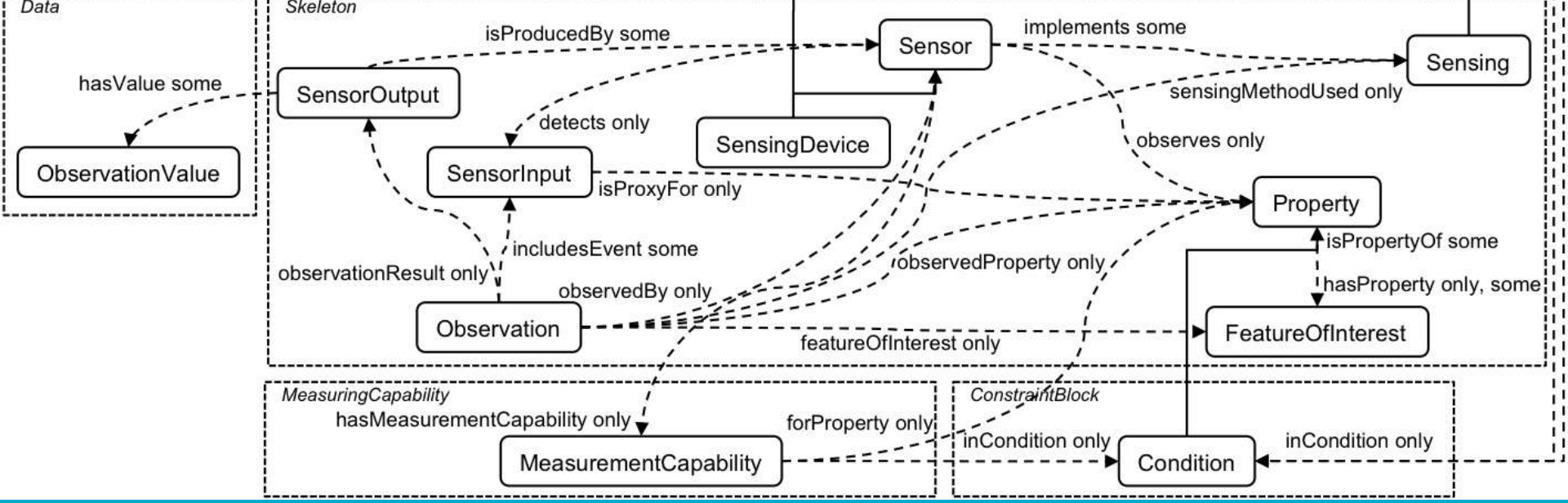


Developing Ontologies for Linked Geospatial Data

Linking Geospatial Data Workshop, London March 6-7 2014

KERRY TAYLOR, CSIRO COMPUTATIONAL INFORMATICS
www.csiro.au

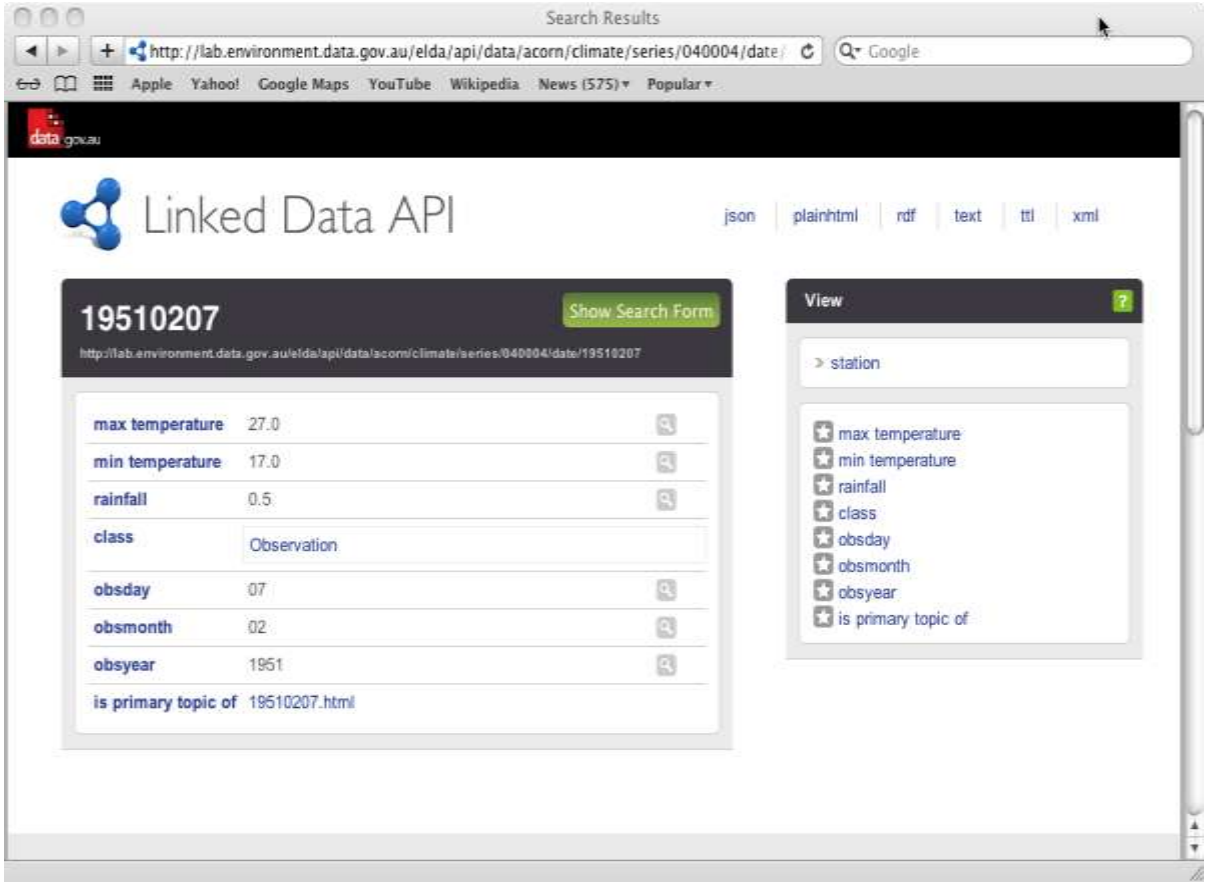
A tale of two ontologies

1. **ACORN-SAT** Long term climate data, published for research and community access on data.gov.au

2. **ad-hoc** modelling the Water Regulations 2008 to help the long tail of data providers to report water data

Ta	Volume of potable water supplied to infrastructure operators (other than recycled water and urban stormwater reuse), expressed in megalitres
Tb	Volume of non-potable water supplied to infrastructure operators (other than recycled water and urban stormwater reuse), expressed in megalitres
Tm	Volume of recycled water supplied to infrastructure operators, expressed in megalitres
Te	Total volume of treated sewage discharges from a sewage discharge point, expressed in megalitres
Tea	Volume of sewage collected, expressed in megalitres
Teb	Volume of sewage supplied to infrastructure operators, expressed in megalitres
Tec	Volume of sewage taken from infrastructure operators, expressed in megalitres
Ted	Volume of sewage taken from sewer mining, expressed in megalitres
Tee	Volume of sewage measured at the inlet to a treatment works, expressed in megalitres
Tef	Volume of treated sewage effluent released from a treatment works, expressed in megalitres
Teg	Volume of sewage collected (other than trade waste sewage), expressed in megalitres
Teh	Volume of trade waste sewage collected, expressed in megalitres
Ts	Total volume of urban stormwater discharges from a stormwater discharge point, expressed in megalitres
Tsa	Volume of urban stormwater supplied to infrastructure operators, expressed in megalitres
Tsb	Volume of urban stormwater taken from infrastructure operators, expressed in megalitres
Tsc	Volume of urban stormwater taken for managed aquifer recharge, expressed in megalitres
Tsd	Volume of urban stormwater used, expressed in megalitres
Tp	Total volume of recycled water supplied, expressed in megalitres
Tpa	Volume of recycled water supplied for residential use, expressed in megalitres
Tph	Volume of recycled water supplied for commercial, municipal or industrial use, expressed in megalitres
Tpc	Volume of recycled water supplied for agricultural irrigation, expressed in megalitres
Tpd	Volume of recycled water supplied for environmental use (other than managed aquifer recharge), expressed in megalitres

Linked Open Data

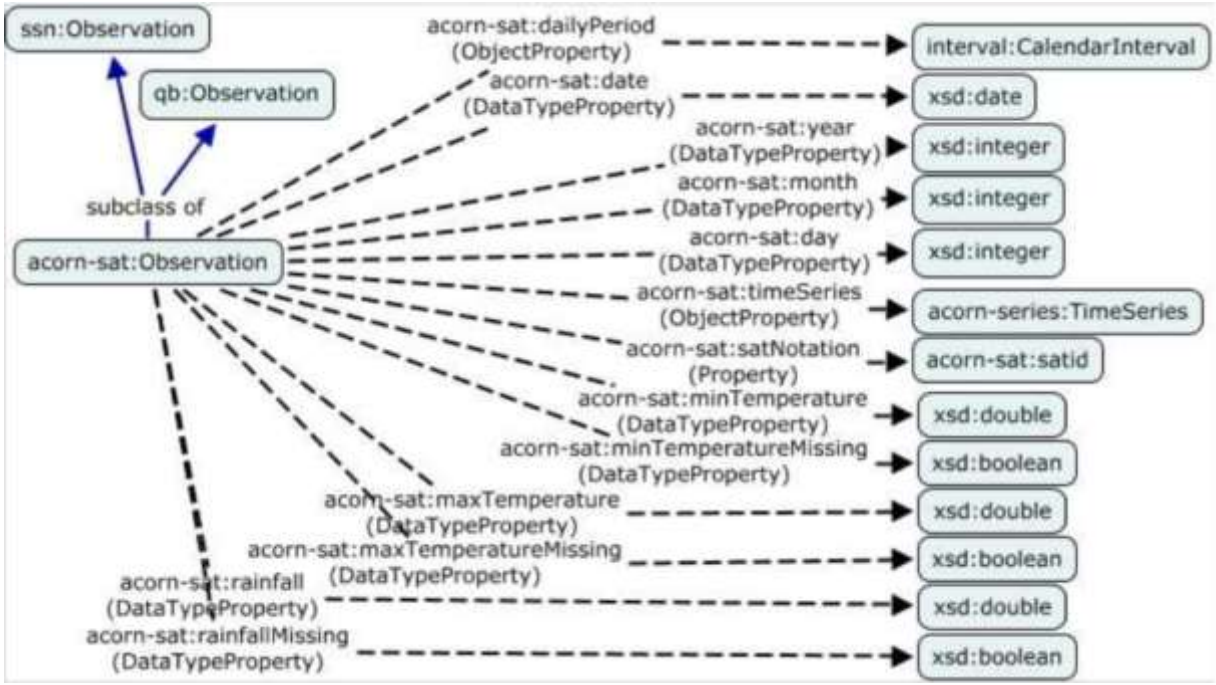


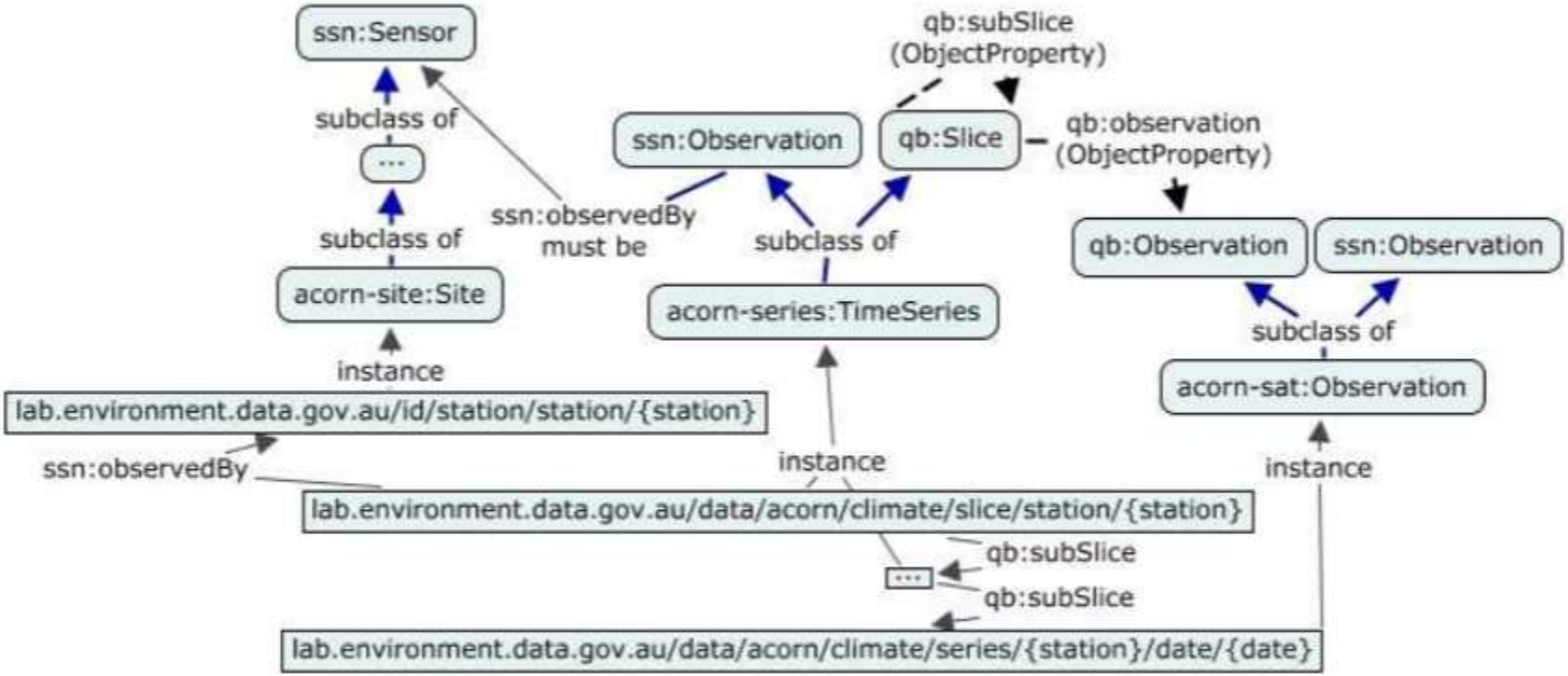
- 100 year climate observations, 61 million triples, Station Metadata included
- Use in any browser via Linked Data API (HTML)
- In Javascript via Linked Data API (JSON)
- In R via SPARQL

Ontology

- Data-driven design and reuse

- Uses W3C SSN (itself loosely derived from O&M) for deployment
- Uses W3C datacube for slices by stations and time periods
- Uses geonames for places




ssn:Sensor
subclass of
...
subclass of
acorn-site:Site
instance
lab.environment.data.gov.au/id/station/station/{station}
ssn:observedBy
ssn:observedBy
must be
qb:subSlice
(ObjectProperty)
ssn:Observation
qb:Slice
qb:observation
(ObjectProperty)
subclass of
acorn-series:TimeSeries
instance
lab.environment.data.gov.au/data/acorn/climate/slice/station/{station}
qb:subSlice
...
qb:subSlice
lab.environment.data.gov.au/data/acorn/climate/series/{station}/date/{date}
qb:Observation
ssn:Observation
subclass of
acorn-sat:Observation
instance

Now

- 6 person-months from CSV and PDF to 5* open linked data
- Australia's first data.gov.au LOD
- Is hosted on Australia's "NCI"
- Epimorpics API (ELDA) on lab.environment.data. gov.au
- RDF/XML on data.gov.au
- Established Australian Government Linked Data Working Group (now developing URI policy)

WDTF ontology

Two reasons for building a water ontology

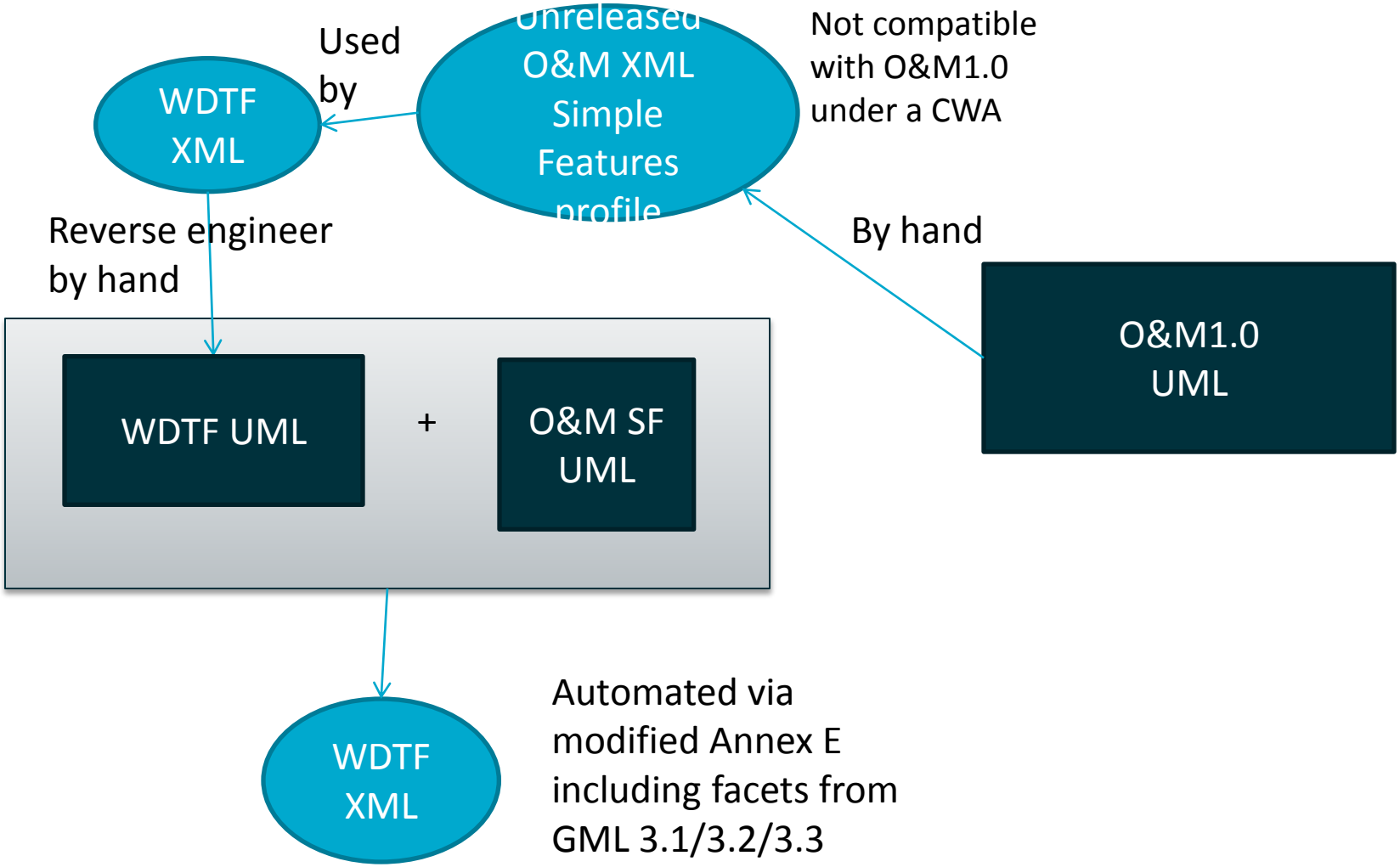
- WDTF is an XML format developed to capture water data from the larger agencies required to report water data to the Bureau of Meteorology.
- Adopted broadly in Aus industry
- GML application, implementing O&M
- Validated by schematron rules
- Later developed & standardised as WaterML2.0 through OGC (then WMO?)
- Experiment with UML to OWL transition

- *Ad hoc data ingestion* is a tool to capture data from the long tail of agencies. Users create mappings from their own spreadsheets to water regulations concepts modelled in an ontology. Reasoner is used but most verification applied outside the ontology <closed-world>

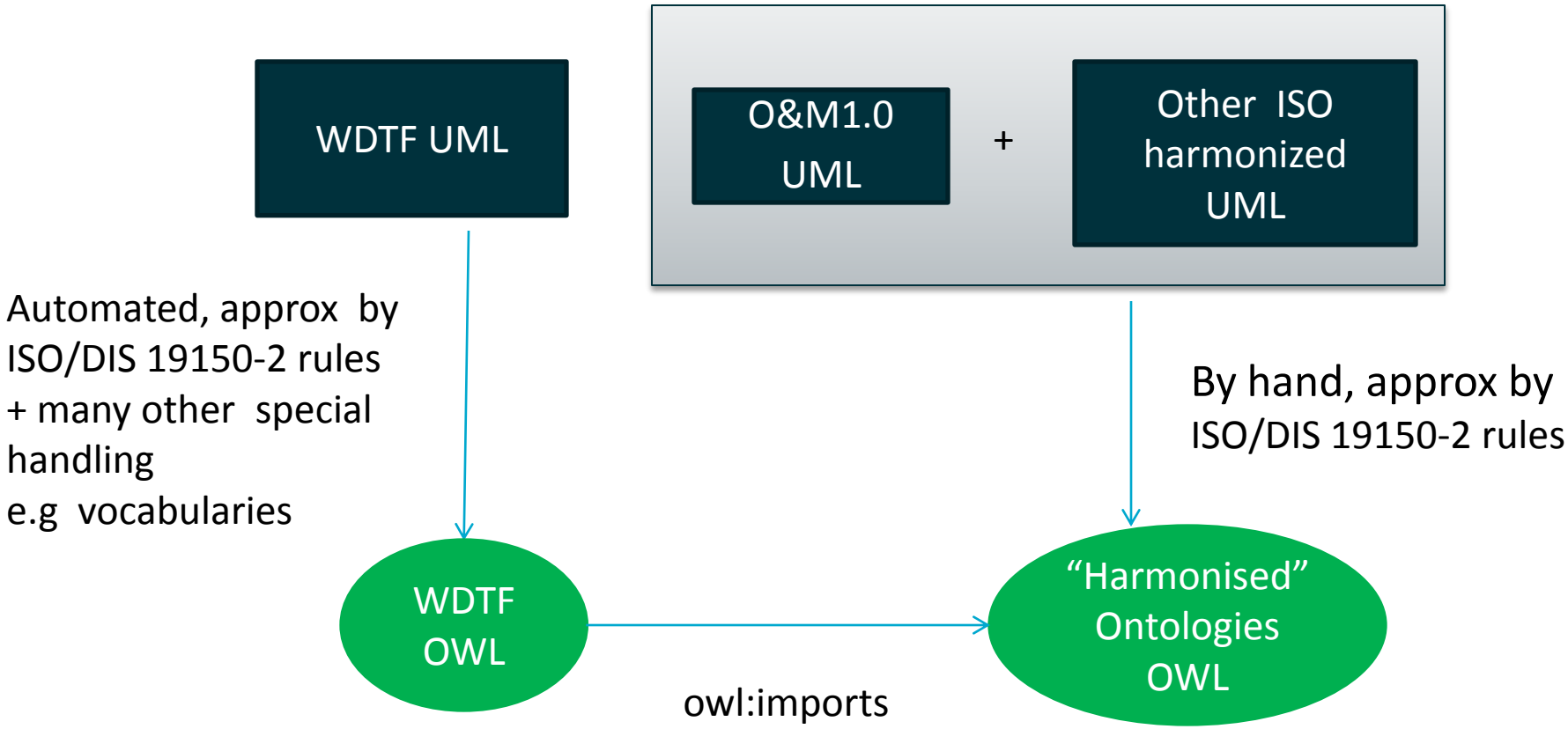
Model Driven Architecture Approach

- Brought these threads together—ISO-inspired one (UML) model to derive them all.
- *Does this sit well with linked data principles and practices?*
- Note that ad-hoc is not specifically a linked data use case, but the ontology is needed for user-communication of data requirements.

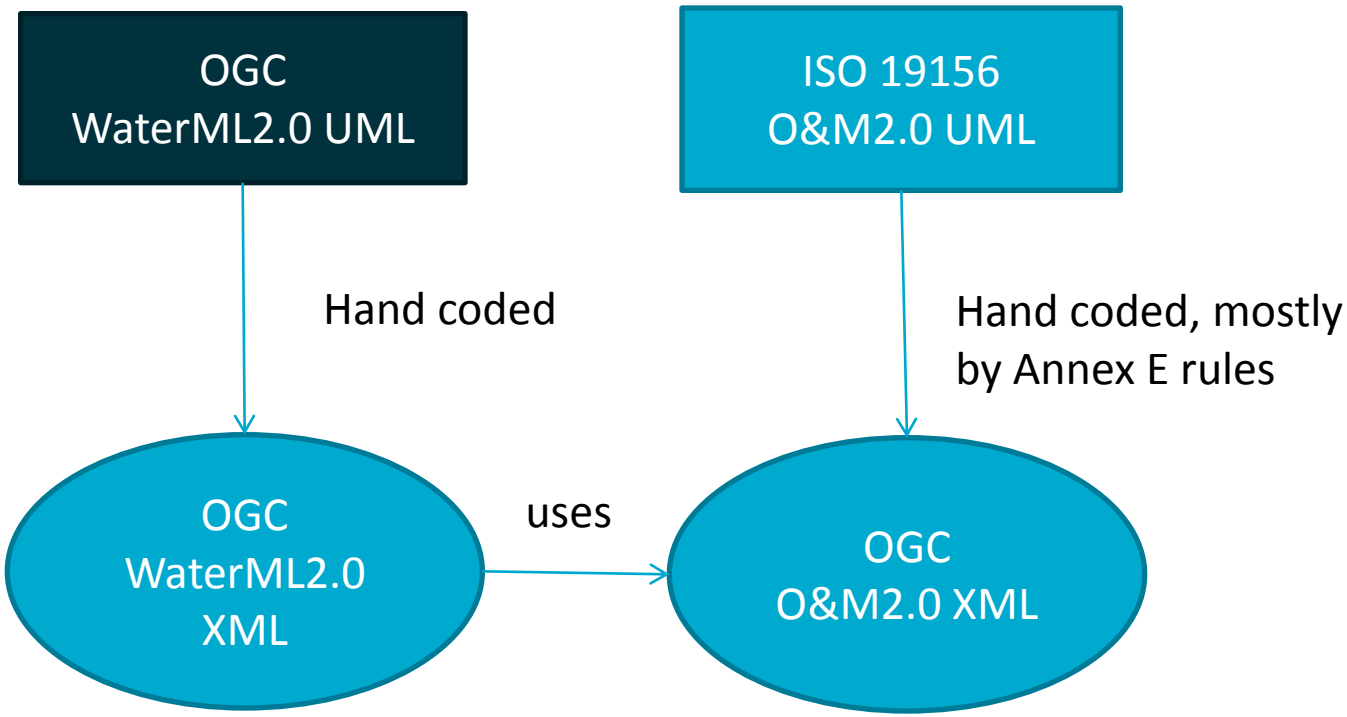
Making WDTF UML and WDTF XML



Making Ad-Hoc WDTF OWL



WaterML2.0 (arose from WDTF)



Hand coding allows simplified schema, supports defaulting and other structural optimisations

Hand coding allows arbitrary XML use in some places instead of only O&M derived XML instances

What went wrong?

Cause

- Multiple inconsistent artefacts at different stages of maturity/stability (is this a permanent feature?)
- Too much “semantics” in UML stereotypes not amenable to generic UML to OWL toolset
- OGC encoding rules (Annex E) informally written and evolving and so ambiguous
- Various handcoding steps and some automation often incompatible
- Modelling languages enforce a “style”; choices may not be intended “conceptual” and therefore automation may be inappropriate
  - “Intended” scoping of property names unclear in UML (now package scope)
  - Modularisation methods(and effect) in UML and OWL are different
  - CWA vs OWA
  - Annotation to carry through the UML encoding practice becomes meaningless in OWL
  - Critical vocabularies not available in UML at all; managed externally

Effect

- WDTF ontology is not OWL-DL and reasoner reports inconsistency
- OWL semantics-driven methods in adhoc tool fail
- Is *not* a simplified domain ontology for user communication (because it looks too much like a GML –driven application schema)
- Will be difficult to maintain

AUTHORS:

Andrew Woolf
Australian Bureau of Meteorology, Canberra Australia
A.Woolf@bom.gov.au

Kerry Taylor, Laurent Lefort, Geoff Squire, Gavin Walker, Yanfeng Shu, David Ratcliffe, Armin Haller , CSIRO Computational Informatics, Canberra, Australia

Simon Cox, CSIRO Land and Water, Melbourne, Australia

Many other CSIRO staff and the OGC and ISO community also contributed to this work

Firstname.Lastname@csiro.au

CSIRO WATER FOR A HEALTHY COUNTRY FLAGSHIP
www.csiro.au