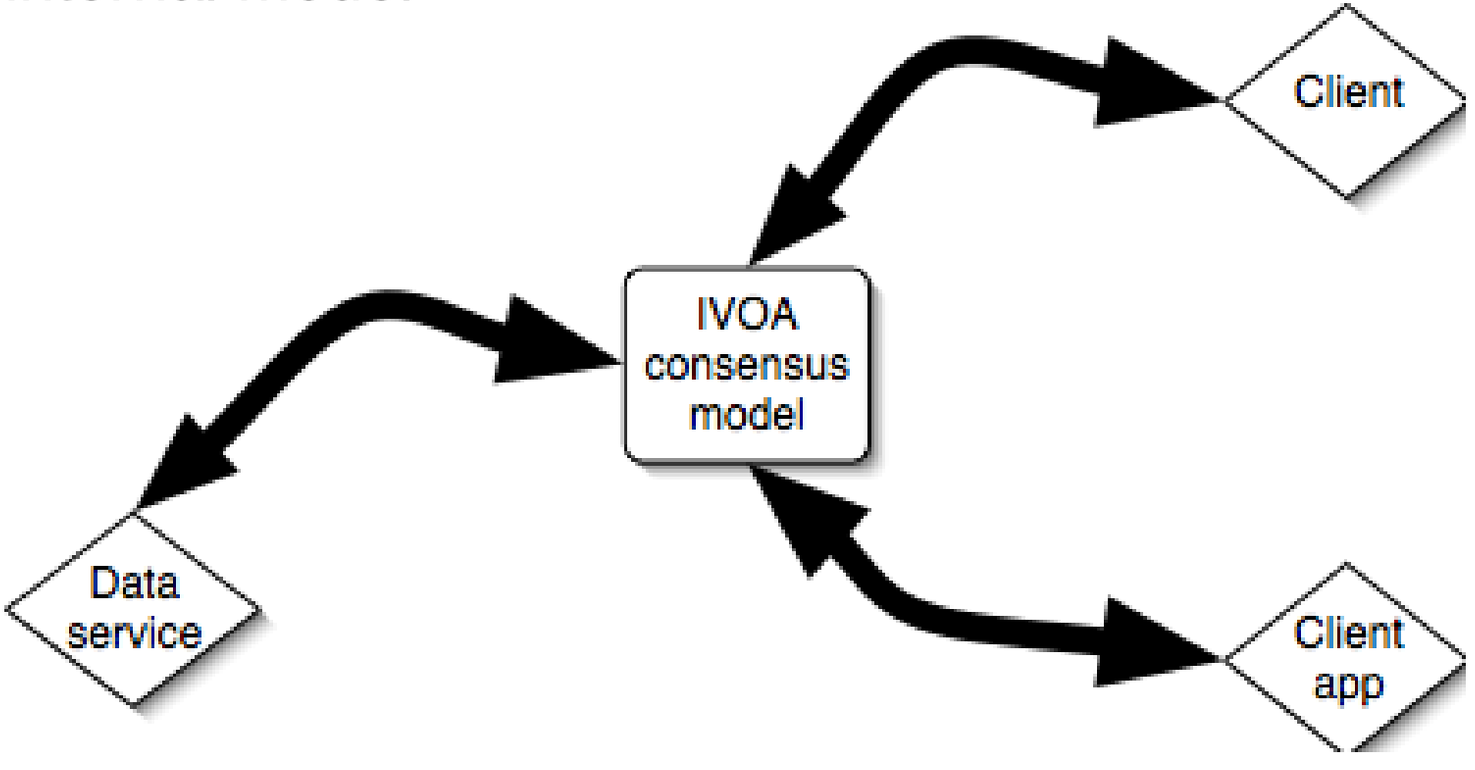
Ontology access to foreign data models

Norman Gray
Starlink / University of Glasgow

2005 May 19 – Kyoto

plan a

Convert data into and out of a consensus model, from an internal model



problems and dangers

- The consensus has to be reached, possibly painfully.
- The compromises will hit different folk in unpredictable ways.
- (VOTable and UCD both started with well-established drafts)
- Big/complete documents are expensive.
- No obvious migration/development route, so v.1 is v.important.

But mostly. . .

problems and dangers

Going beyond the consensus will be *hard/expensive* – ie, non-interoperable – and change in time is equivalent to this.

(this is of course why we need a standard in the first place)

Are lots of translator services the only solution? Yes, if an n^2 solution is the only alternative.

this is a good plan. . .

. . . but there is a danger that the DM effort could converge on a product which is of real use in a good range of situations, but which will be either too complicated or too simple for some.

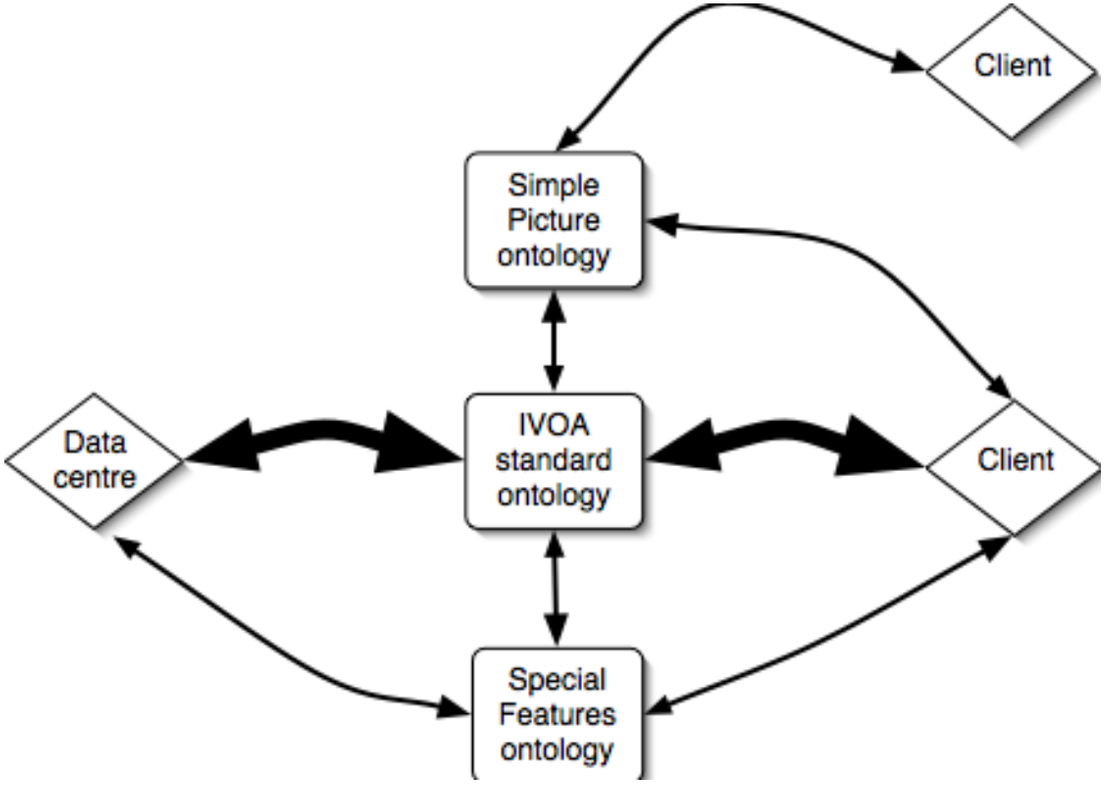
- HDX/NDX: the application wants a simpler model.
- ORAC-DR uses metadata to drive processing, and thus needs a model richer than the IVOA DM can reasonably be.
- The bigger the IVOA-DM's range of applicability, the longer and more painful will be the process (and the bigger will be the documentation).

plan b

Plan B is to give multiple DMs (approx) equal status, and handle that by making it easy for applications to ‘understand’ foreign DMs.

- Here ‘multiple’ is not necessarily ‘many’, but it’s more than we ought to expect a single application to implement natively.
- Range from simple to complicated; and widely to rarely implemented
- . . . well-standardised to casual?

that is. . .


Client
Simple Picture ontology
Data centre
IVOA standard ontology
Client
Special Features ontology

using ontologies

Communicate ‘understanding’ by linking between DMs:

- ‘the content of the first FITS extension has the same meaning as the content of the HDX Image object’
- ‘the content of the third database column is identical to the telescope pointing RA’
- ‘the telescope pointing RA is usable as the UCD RA of the image centre, in the context of making Pretty Pictures

to-wit to-woo

OWL is designed for making these ontologies and declaring the relationships of their properties to each other and to those in other ontologies.

There are systems in existence which will do the inferencing required to answer questions like. . .

> ‘I know about HDX and I know about UCD: I’ve been given this database metadata and I want something like a UCD RA: which database column should I retrieve?’.

the payoff

- Individuals, from data providers to application authors, get to *choose* which of several data models are most valuable to them, *without* sacrificing interoperability.
- Thus, it's now safe to go beyond the consensus.
- Merging ontologies is easy/easier.
- Versioning/development becomes easier.
- Rigourously interoperable, and as standardisable as you want.

practicalities

Those systems aren't currently easy to build into applications.

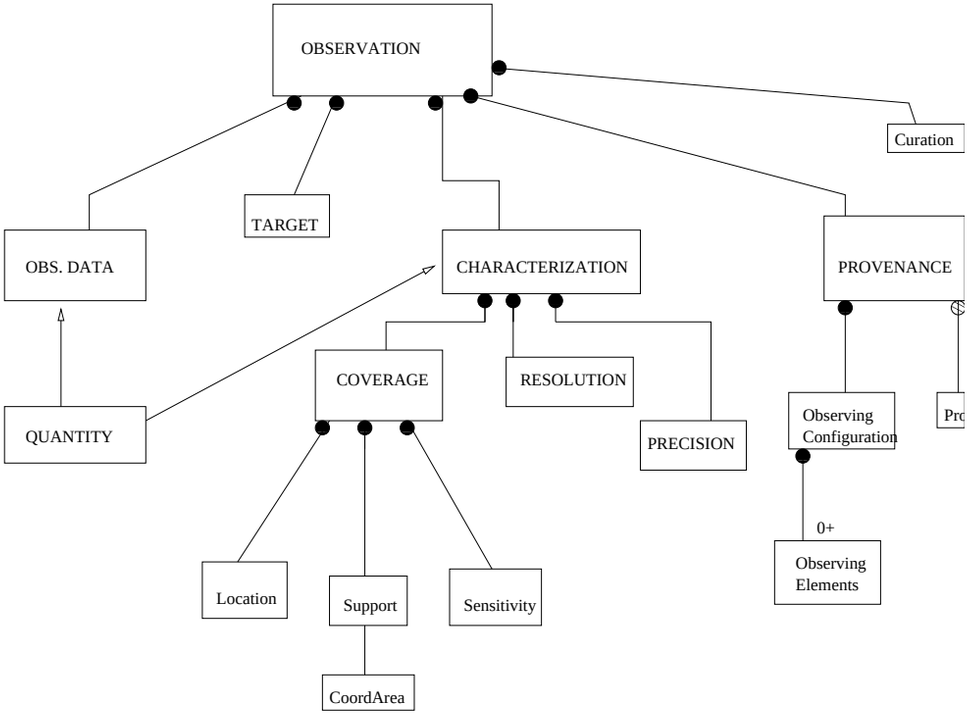
- Need to work out what level of standardisation is required for the OWL ontologies (indeed if *any* extra is required),
- . . . turn the inferencing step into a practical library or service,
- . . . discover what would it take to make this technology as a whole available to non-specialist application and data-service developers?

How practical? How many DMs are enough? Or too many?

demo: alberto's utypes

- From Alberto's UTYPE note
- Assert: *Observation* has the *hasCharacterisation* property, with range *AlbertoCharacterisation*
- Assert: An *ImageObservation* necessarily has a *hasCharacterisation* property, with restrictions on the values; also this is a subclass of *sim:ImageData*.
- Classify: nothing particularly exciting, except that the resulting tree has *ImageObservation* under *sim:ImageData*

demo: dm for obs 0.23 figure 4



Model only *Observation*, *ObservationData*, and *Provenance*.

the classifier

- We imported the *alberto* ontology, and mentioned one of its properties.
- The *Observation* class has an *alberto:hasCharacterisation* property; ⇒ it must be a type of `alberto:Observation`.
- We constrained the types of *alberto:hasCharacterisation* our *ImageObservation* has; ⇒ this must be a type of *alberto:ImageObservation*, and thus a type of *sim:SimpleImage* – a class we'd never heard of before.

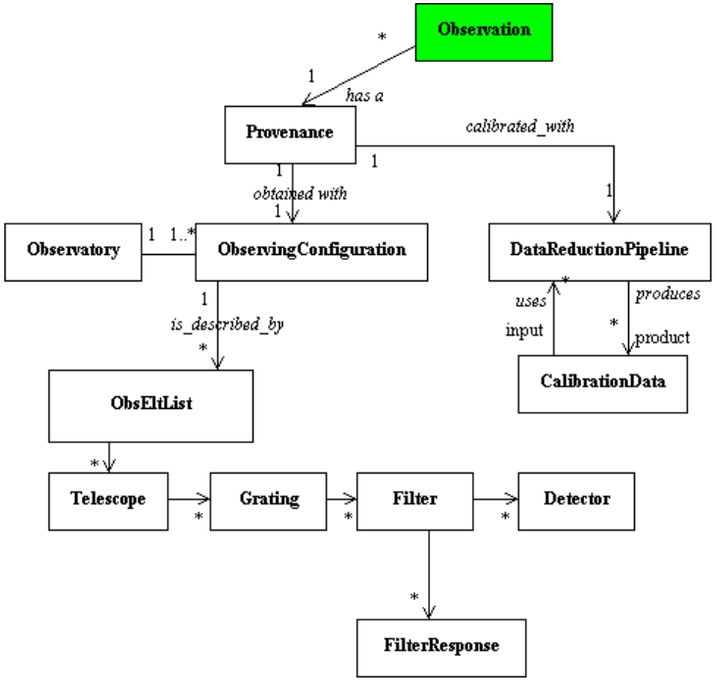
the classifier (2)

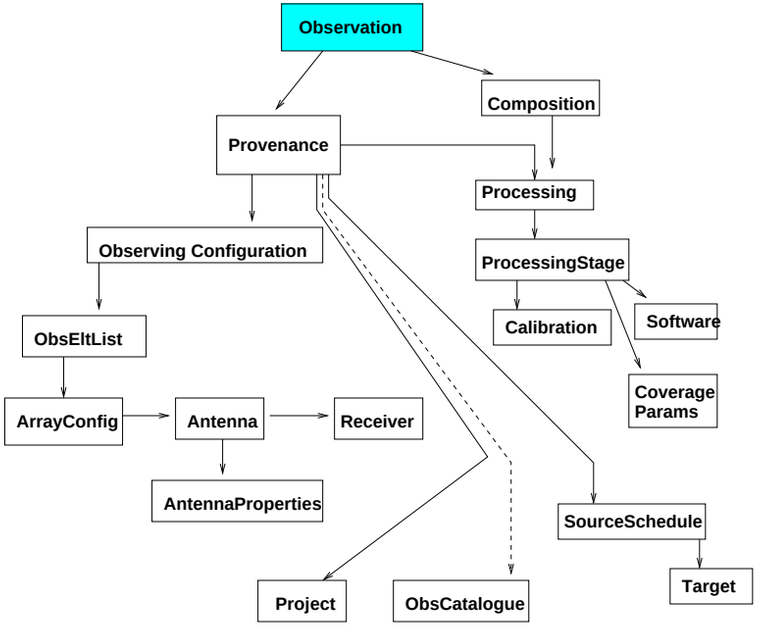
- The class *SpectralImageObservation* is inconsistent – oops!

. . . benefits

- You are looking for a *sim:SimpleImage*? Find an *ImageObservation* & ignore *SpectralObservation*.
- Use ontology compilation to catch merging bugs and conflicts.
- The inferred ‘subsumption hierarchy’ is already complicated – be glad you don’t have to work it out by hand.

provenance objects





Which do *you* prefer?

another community's experiences: bioinformatics

- Bioinformatics: the Gene Ontology. Like some others, very pragmatic, usually graphs or taxonomies (computer scientists find them rather dull).

- The inferencing is typically not 'live', but used to 'compile' a set of rules into a simple tree-based ontology.

___how to make a good ontology/data model

- The wider community has to be involved $\Rightarrow$ clear and limited goals
- Simple result
- Continuous evolution (the reasoner can help), active curation
- Early use
- No Knowledge Engineers (visibly) involved

http://www.astro.gla.ac.uk/users/norman/

contact

- http://www.starlink.ac.uk
- http://www.astro.gla.ac.uk/users/norman/
- norman@astro.gla.ac.uk