

Intelligent access to foreign data models

Norman Gray
Starlink / University of Glasgow

2004 November 18

Starlink

Convert data into and out of a consensus model

- | ...to and from an internal model (explicit or not)
- | For example, TopCat gains power by using plugins implementing an internal model in terms of external models (explicit rather than unusual)
- | A Good Plan

Starlink

problems

- The consensus has to be reached, possibly painfully.
- The compromises will hit different folk in unpredictable ways.
- Such a consensus has been reached for VOTable and more-or-less for UCD, but in both cases there was a fairly well-established draft.
- The IVOA data models group, starting without the first advantage, has taken longer than anyone feared; and I'm pessimistic in a way I'll explain.

Starlink

_____ don't all shout at once

This is still a Good Plan.

- | The VOTable and UCD standards allow things to move forward now, and their role will continue to be important even if they don't completely satisfy the most optimistic goals
- | ...that is, for VOTable, lots of structure and semantics; for UCD, processing-driving semantics and flexible description.

Starlink

_____my guess is...

... that the DM effort will 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 a whole host of other uses.

- 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 (plus documentation cost).

Starlink

_____ (stc is special)

All of this doesn't apply to STC, since

- | everyone happens to have more-or-less the same model of what's going on,
- | though they'd like it to be as simple as possible, they know it's more complicated than they necessarily remember,
- | and they're quite happy for Arnold to get stuck into the Explanatory Supplement on their behalf.

Starlink

Plan B is to give multiple DMs 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
- | ... widely to rarely implemented
- | ... well-standardised to casual.

Starlink

_____using ontologies

Communicate 'understanding' via statements like

- '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 (that is, unless you find something else which is identical to the UCD RA of the image centre)'

Starlink

_____mapping ontologies

That is, we make *links between* data models, rather than defining a data model in isolation.

Just link the foreign model to a model your application already understands.

Starlink

_____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?’.

Starlink

_____where do these assertions come from?

The OWL (or whatever) assertions linking the database metadata could either:

- | come with the metadata,
- | are local/private to the application,
- | or are a thirdparty mapping of the database's declared model to a variety of other more or less well known models.

Transitivity; inferencing; chinese whispers.

Starlink

the payoff

Individuals, from data providers to application authors, get to choose which of several data models are most valuable to them. And they can use that model, without sacrificing interoperability.

The cost : you have to use an extra step to interpret a foreign model;

The payoff : you can read that other model without both you and the data supplier having to do a private translation to and from the consensus model.

Or in other words: you write your TopCap plugin automatically.

Starlink

practicalities

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

- I want 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?

Starlink

| <http://www.starlink.ac.uk>

| <http://www.astro.gla.ac.uk/users/norman/>

| norman@astro.gla.ac.uk