Metadata and Semantics in the Virtual Observatory

Norman Gray

EuroVO-TECH / AstroGrid / Leicester / Glasgow, UK

Glasgow DCS, IR group, 2008 April 21

logos

A word from our sponsors…

rdf

Resource Description *Framework*

- All the world is triples, consisting of resources named by URIs (http:...)
- ...which have properties whose values are resources or literals.
- RDF/RDFS/OWL describe these using rdf:type, rdfs:subClassOf, owl:symmetricProperty, and so on.

There is an analogy with XML Schemas, *but it is a loose one* – they're not addressing the same problem. Same for O-O.

rdf/semweb/rdb/xml – strengths

- RDF/OWL/Reasoning now largely stable (though The Semantic Web will remain Vision)
- Using the architectural principles which helped HTTP take over the internet. Open and flexible.
- RDB to XML to RDF – spectrum of strengths. XML is more natural than RDF where the information density is high, and the information regular or highly constrained; RDF/SW is natural for incomplete or ragged data.

rdf gives you reasoning

```
<http://x> a ns:SecondaryEducationContentLevel.
ns:SecondaryEducationContentLevel
  rdfs:subClassOf
    ns:SchoolContentLevel.
```

Thus `http://x` is School Content Level, too.

Or…

```
<http://x> ns:emailAddress <mailto:foo@example.org>.
<http://y> ns:emailAddress <mailto:foo@example.org>.
ns:emailAddress a owl:InverseFunctionalProperty.
```

implies `<http://x>` and `<http://y>` are the same entity.

so…

- Add transitive, (inverse) functional & symmetric properties, subclass/subproperty relations, and you magnify what you say.
- The win is that the reasoner can expand the set of assertions in your knowledgebase, by drawing all possible conclusions.
- Then add derived types, annotations, easy extension and more.
- Multiple syntaxes: N3/Turtle and RDF/XML.

db volume

Triplestores are for bulk instances, and trade off volume/speed vs. expressiveness: fast RDFS reasoning vs. slow OWL reasoning.

- Expressiveness: RDFS, OWL DLP, OWLIM, OWL Lite, OWL DL, SWRL, OWL Full.
- This year, 10^8 triples is A Lot, and 10^9 is Expensive.
- Hybrid solution: offline OWL reasoning ‘compiled’ to bulk RDFS assertions.

sparql

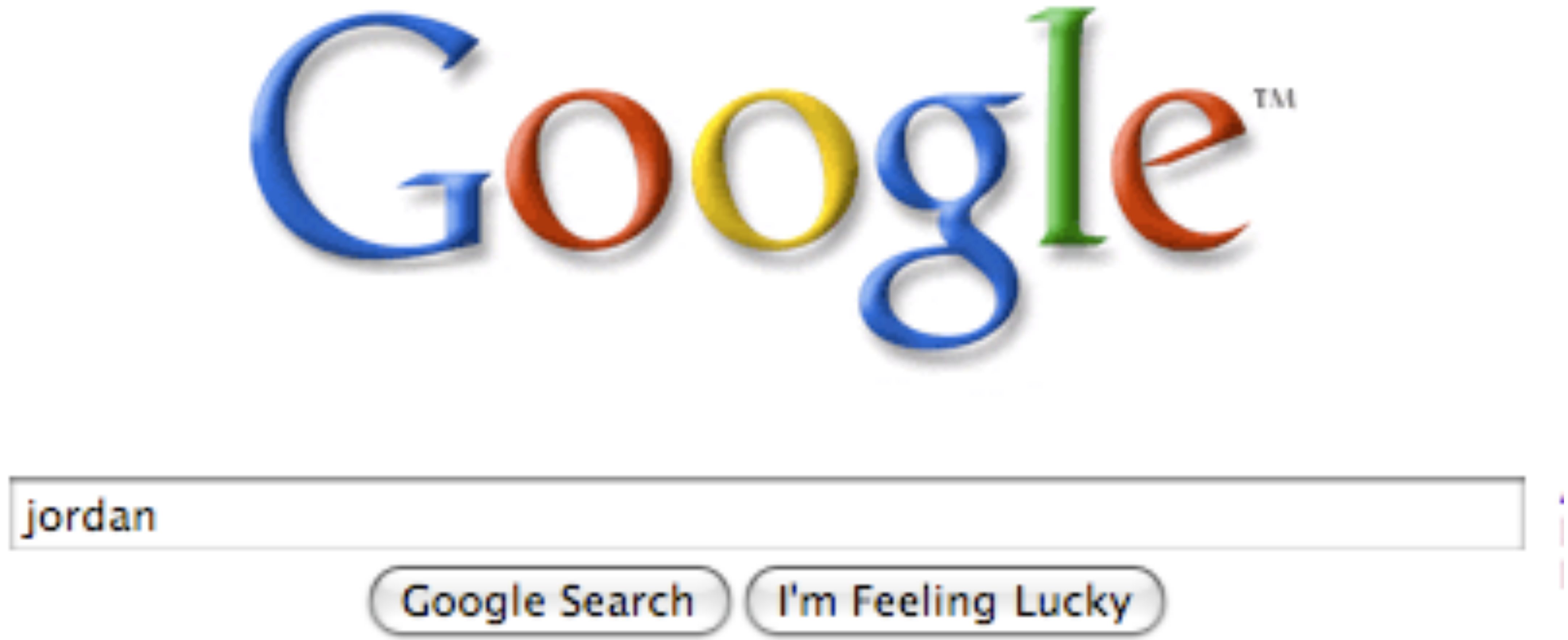
queries, i

```
prefix vor: <http://www.ivoa.net/xml/VOResource/v1.0#>
prefix me: <http://example.org/norman#>

select ?r ?title
where {
  ?r vor:title ?title
  ?r a me:ResearchAtlas.
}
```

queries, ii

```
prefix vor: <http://www.ivoa.net/xml/VOResource/v1.0#>
prefix sia: <http://www.ivoa.net/xml/SIA/v1.0#>

select ?r ?t
where {
  ?r vor:capability ?cap.
  ?cap [ sia:imageServiceType [ a sia:ImageServiceTypeAtlas ] ].
  ?r vor:content [ vor:contentLevel [ a vor:ResearchContentLevel ] ].
  ?r vor:identifier [ vor:authorityID ?authid].
  FILTER REGEX(?authid, "\.ca$") .
  ?r vor:title ?t.
}
```

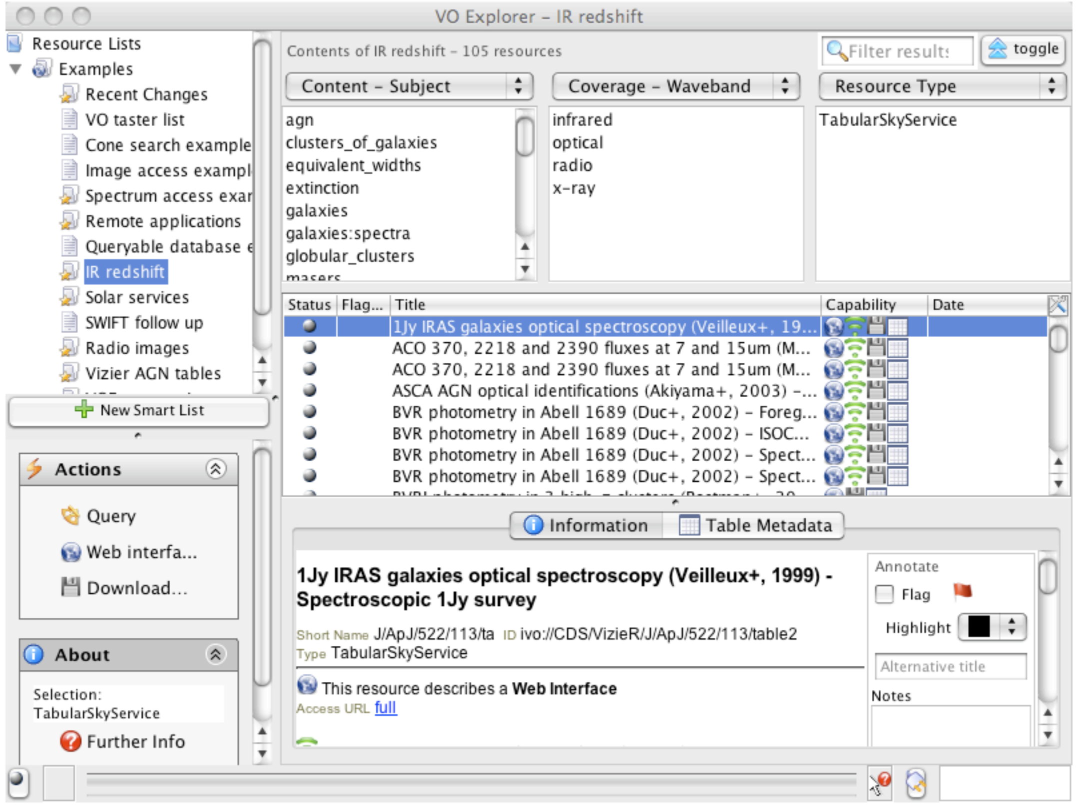
metadata and the semantic web



jordan

Google is clever, but it’s not psychic

voexplorer



VO Explorer – IR redshift

- Resource Lists
- Examples
  - Recent Changes
  - VO taster list
  - Cone search example
  - Image access exampl
  - Spectrum access exar
  - Remote applications
  - Queryable database e
  - IR redshift
  - Solar services
  - SWIFT follow up
  - Radio images
  - Vizier AGN tables

New Smart List

Actions
- Query
- Web interfa...
- Download...

About

Selection:
TabularSkyService

Further Info

Contents of IR redshift – 105 resources

Filter results | toggle

Content – Subject	Coverage – Waveband	Resource Type
agn	infrared	TabularSkyService
clusters_of_galaxies	optical	
equivalent_widths	radio	
extinction	x-ray	
galaxies		
galaxies:spectra		
globular_clusters		
masers		

Status	Flag...	Title	Capability	Date
●		1Jy IRAS galaxies optical spectroscopy (Veilleux+, 19...		
●		ACO 370, 2218 and 2390 fluxes at 7 and 15um (M...		
●		ACO 370, 2218 and 2390 fluxes at 7 and 15um (M...		
●		ASCA AGN optical identifications (Akiyama+, 2003) –...		
●		BVR photometry in Abell 1689 (Duc+, 2002) – Foreg...		
●		BVR photometry in Abell 1689 (Duc+, 2002) – ISOC...		
●		BVR photometry in Abell 1689 (Duc+, 2002) – Spect...		
●		BVR photometry in Abell 1689 (Duc+, 2002) – Spect...		

Information | Table Metadata

1Jy IRAS galaxies optical spectroscopy (Veilleux+, 1999) - Spectroscopic 1Jy survey

Short Name J/ApJ/522/113/ta ID ivo://CDS/VizieR/J/ApJ/522/113/table2
Type TabularSkyService

This resource describes a **Web Interface**
Access URL full

Annotate

□ Flag

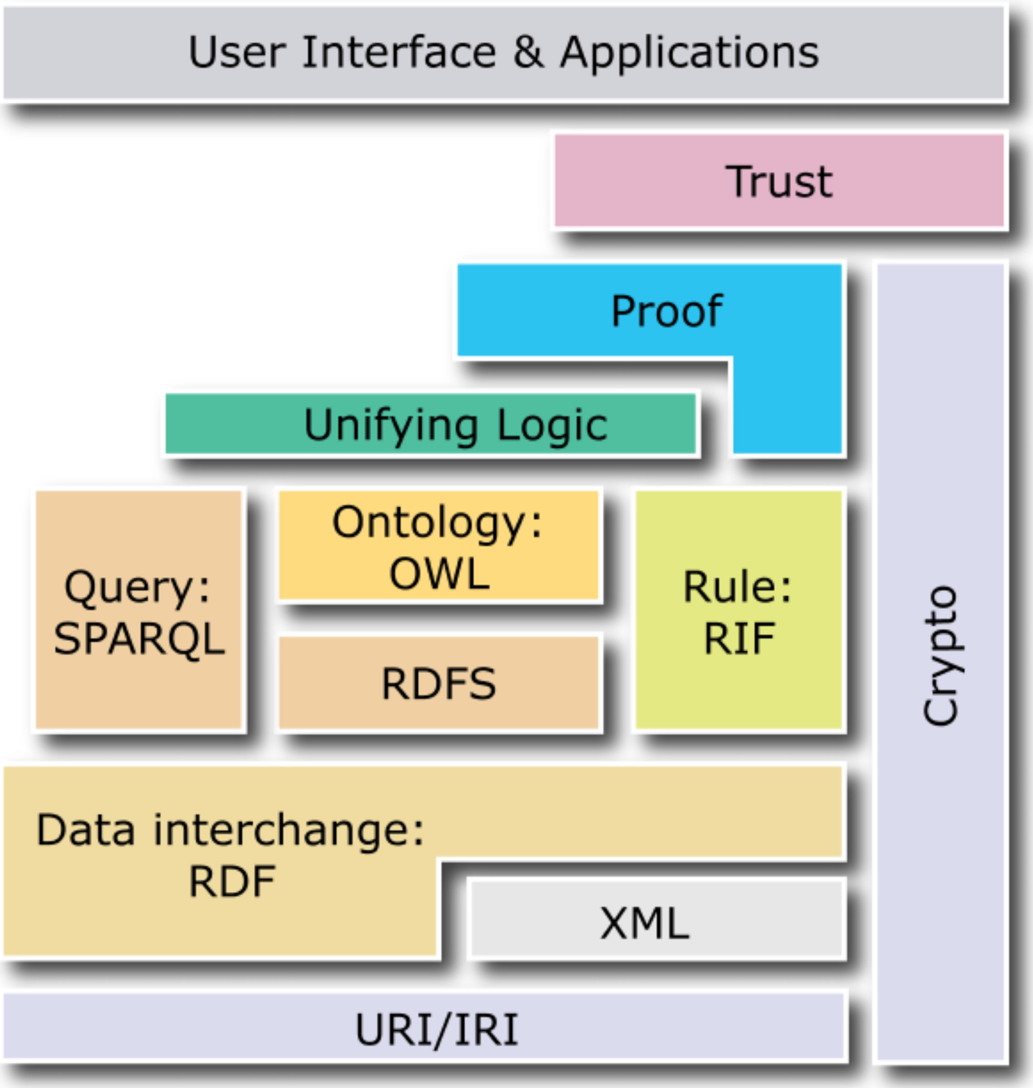
Highlight ■

Alternative title

Notes

Faceted browsing – clever string searching

the semantic web



but where’s the metadata going to come from?

asking for `<#compactObject>` and getting `<#blackHole>` is easy

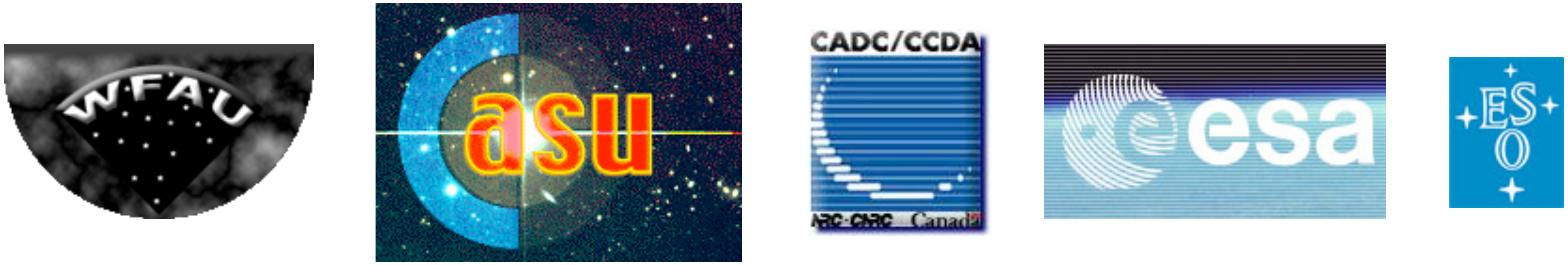
from the users?

ebay + “plam pilot”

http://www.well.com/~doctorow/metacrap.htm

- self-interest helps, **if the UI is right**
- ...so does cunning
- but this is still playing tricks with strings, and astronomy can surely do better

from the data sources?



- depends on the UI, again
- ...but they can be embarrassed

astronomy
& the virtual observatory

The Virtual Observatory

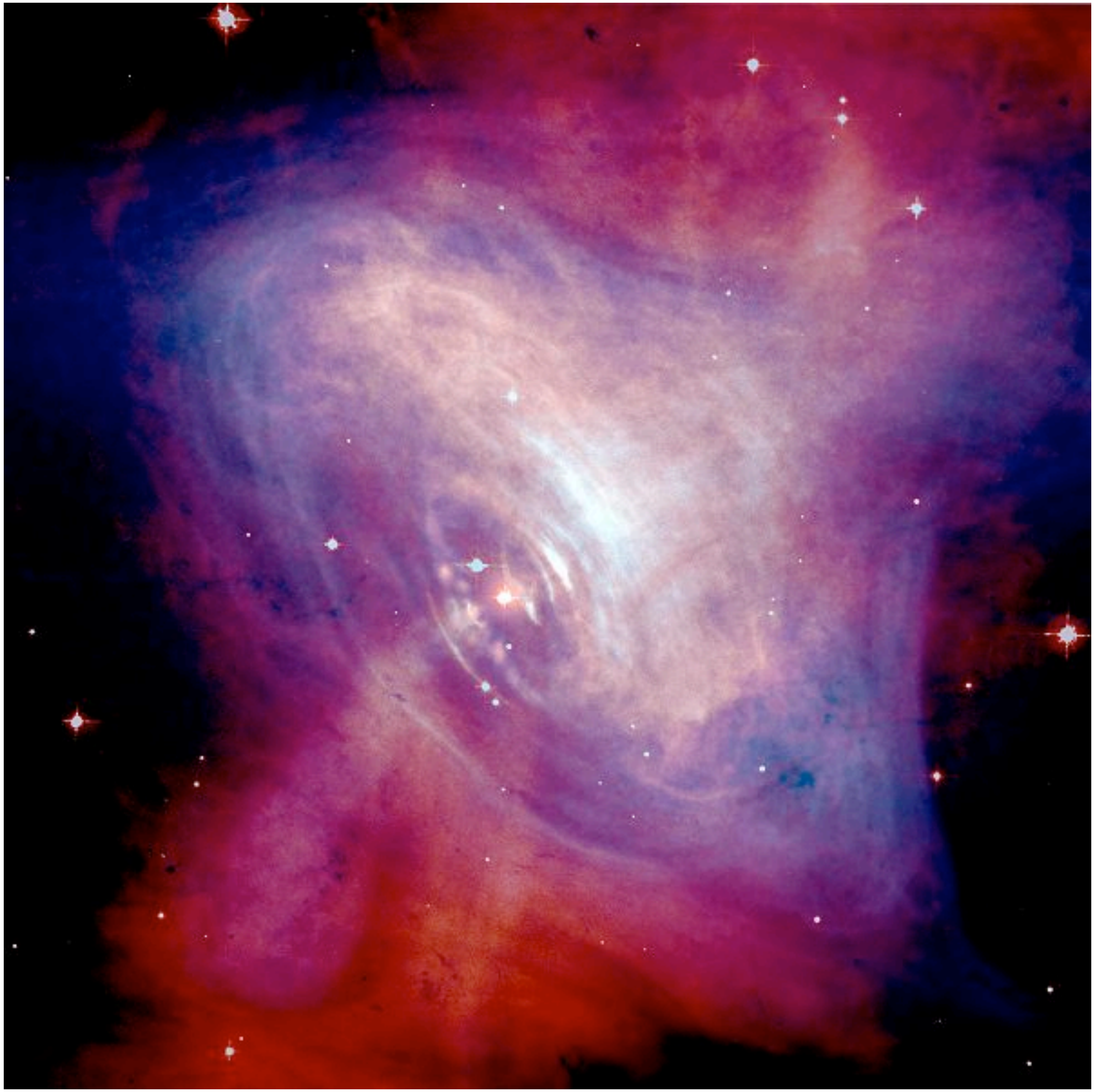
- Observing is expensive
- But observatories have archives
- ...so perhaps the image you want already exists.
- Make multiwavelength astronomy easy
- ...and save all that tedious trolling off to Hawai‘i

vo problems

- How do I find data?
- Once I've found it, how do I ask for it?
- Once I've got it, what do I do with it?

scenario

- Aloysius the Astronomer's favourite object is the Crab – what does it look like in radio and in X-rays?
- `<clickety-click...>`
- It's also known as M1, NGC 1952, Cul 0531+219, ...
- Aha!



scenario

Mathilda is reading a paper online. She drags the paper into VOExplorer and asks for ‘more like this’. VOExplorer calls out to a service which finds the AOIM and Simbad equivalences of the A&A keywords, and uses the former to query a suitable service to find some pretty pictures (APOD), and the latter to query Simbad, presenting the two lists to Mathilda. There aren’t many pretty pictures, so Mathilda asks to expand the search, and VOExplorer asks for pretty pictures corresponding to a more general term, found either directly in the AOIM vocabulary, or finding a more general Simbad term and finding the AOIM equivalent of that. The Simbad query, on the other hand, has produced far too many hits, so VOExplorer looks down the tree of Simbad terms which are ‘narrower’, and asks ‘you were looking for compact objects: do you mean black holes, quasars, or...?’ Once she has established a suitable keyword or keywords, she can make a queries using the equivalent terms in whichever vocabularies the registry or VOEvent keywords are drawn from. She finds some heterodyne observations, but she’s an X-ray person, so is a bit vague, and curious, about just what that is – but oooh, there’s a link to DBpedia/Wikipedia, so she goes there on the off-chance the article is decent.

projects

- AstroGrid – originally an e-science project, now STFC funded in AG3, for a further three years (maybe)
- Euro-VO – EU-funded
- NVO – US project
- JVO, GAVO, CVO, India, ...
- The IVOA

Prepare to be assimilated!

the ivoa is…

- AstroGrid, Euro-VO, NVO … and everyone who turns up to the meetings
- A standards process modelled on W3C
- Cue bickering, politics, pig-headedness, compromises…
- …and standards folk are happy to sign up to
- Standards cover data transport, access, security, modelling, metadata and more.

vo metadata

- FITS headers: the workhorse
- Registries: describe data and services; rich vs. sparse registration
- UCDs, UTypes, vocabularies: variously principled, variously structured
- Data models: source of much argument
- VOTable: XML data transport

there is metadata there...

```
SIMPLE  =                    T / file does conform to FITS standard
BITPIX  =                   16 / number of bits per data pixel
NAXIS   =                    2 / number of data axes
NAXIS1  =                  512 / length of data axis   1
NAXIS2  =                  512 / length of data axis   2
TELESCOP= 'ROSAT   '           / mission name
OBS_MODE= 'POINTING'           / obs mode: POINTING,SLEW, OR SCAN
PROC_SYS= 'SASS7_3_0'          / Processing system
PROCDATE= '26-AUG-1994 15:50:38' / SASS SEQ processing start date
CHECKSUM= 'ZCL8gBJ8ZBJ8dBJ8'   / HDU checksum updated on 02/02/96
OBJECT  = 'NGC 1275'           / name of object
OBS_ID  = 'WG800591H-1.N1'     / observation ID
OBSERVER= 'VOGES, DR, WOLFGANG,H.' / PI name
DATE-OBS= '05/08/94'           / UT date of obs start (DD/MM/YY)
TIME-OBS= '01:02:20.000'       / UT time of obs start (HH:MM:SS)
RADECSYS= 'FK5     '           / WCS for this file
CTYPE1  = 'RA---TAN'           / Axis type for dim 1 (e.g. RA---TAN)
CTYPE2  = 'DEC--TAN'           / Axis type for dim 2 (e.g. DEC--TAN)
CRVAL1  =       4.99500008E+01 / Sky coord of 1st axis (deg)
CRVAL2  =       4.15099983E+01 / Sky coord of 2nd axis (deg)
```

vo registry

```
<?xml version="1.0" encoding="UTF-8"?>
<vor:Resource updated="2006-11-16"
              xsi:type="tdb:TabularDB">
  <vrx:title>USNO-B</vrx:title>
  <vrx:identifier>ivo://roe.ac.uk/DSA_USNOB/TDB</vrx:identifier>
  <vrx:curation>
    <vrx:publisher>Royal Observatory Edinburgh</vrx:publisher>
    <vrx:contact>
      <vrx:name>Martin Hill</vrx:name>
      <vrx:email>mch@roe.ac.uk</vrx:email>
    </vrx:contact>
  </vrx:curation>
  <vrx:content>
    <vrx:description/>
    <vrx:referenceURL>http://astrogrid.roe.ac.uk:8080/pal-usnob/</vrx:referenceURL>
    <vrx:type>Catalog</vrx:type>
  </vrx:content>
  [...]
</vor:Resource>
```

vo registry

- Around 15000 registry entries in total
- 566/761 registry entries have a non-empty `<subject>`
- Keywords like ‘AGN’, ‘Survey source’, ‘Galaxy cluster’ & ‘cluster of galaxies’, and ‘Binaries:cataclysmic’
- Not too bad, but it could be better

projects

explicator

- `http://explicator.dcs.gla.ac.uk/` : Glasgow (I Ounis & A Gray) & Leicester (N Gray)
- EPSRC-funded project to investigate peer-to-peer ontology mediation in the VO and in HEP
- Allow data centres to distribute data using their own vocabularies, rather than an expensively standardised consensus one
- Avoid losing interoperability by supporting the centres in mutually 'explaining' their vocabularies, doing the required reasoning in a lightweight way.

vocabularies

- A&A keywords (311), AOIM (208), IAU Thesaurus (2951) – forthcoming IVOA standard, using SKOS
- Link keywords together, exploit hierarchy, and help applications help users
- SKOS encapsulates Library experience (operational semantics, not logic)
- Vocabularies target humans (and so support UIs), but provide a bridge to formal concepts, and ontologies
- Map between vocabularies

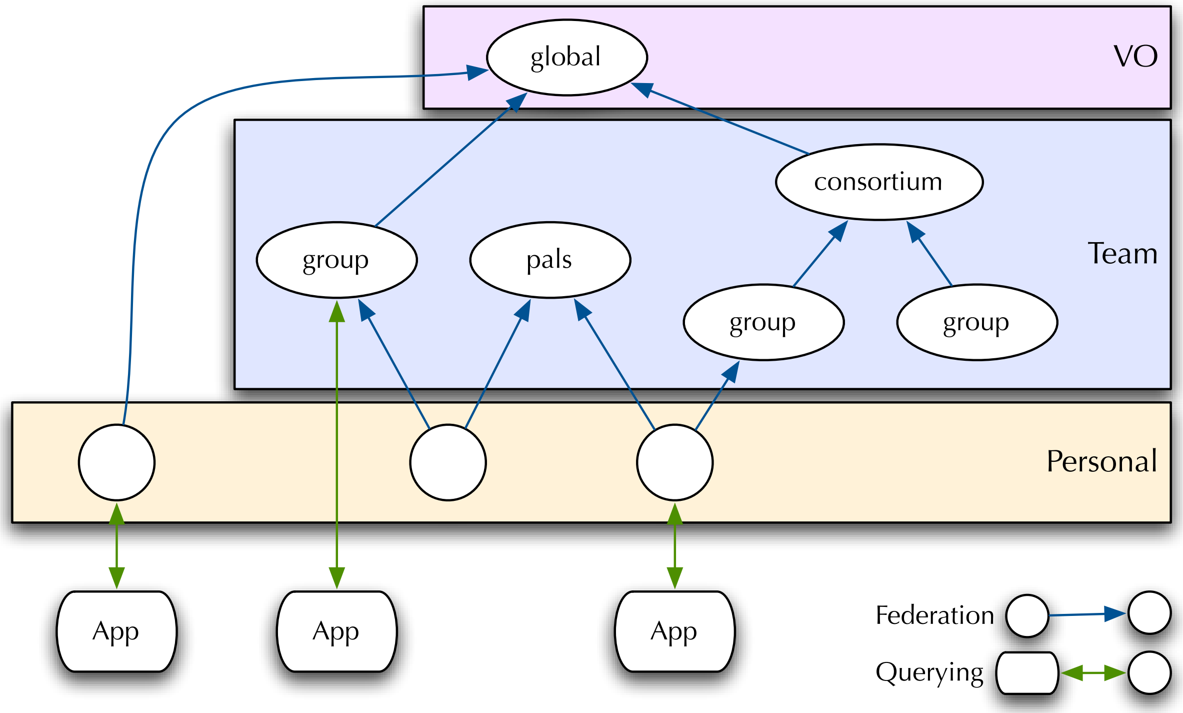
JISC

SKUA

- http://myskua.org
- federated, astronomically aware shared annotation services
- plus Spacebook: social software, linked to myExperiment (Tony Linde)

SKUA architecture

- To service applications, not users
- Apps query 'local' store
- ...which delegates
- Users share material with group and collaborators



astronomical metadata

- Lots of astronomical metadata is available, but it's somewhat ragged
- Projects Explicator and SKUA exploit metadata for interoperability
- Astronomy should make a good SW demo, exploring usability