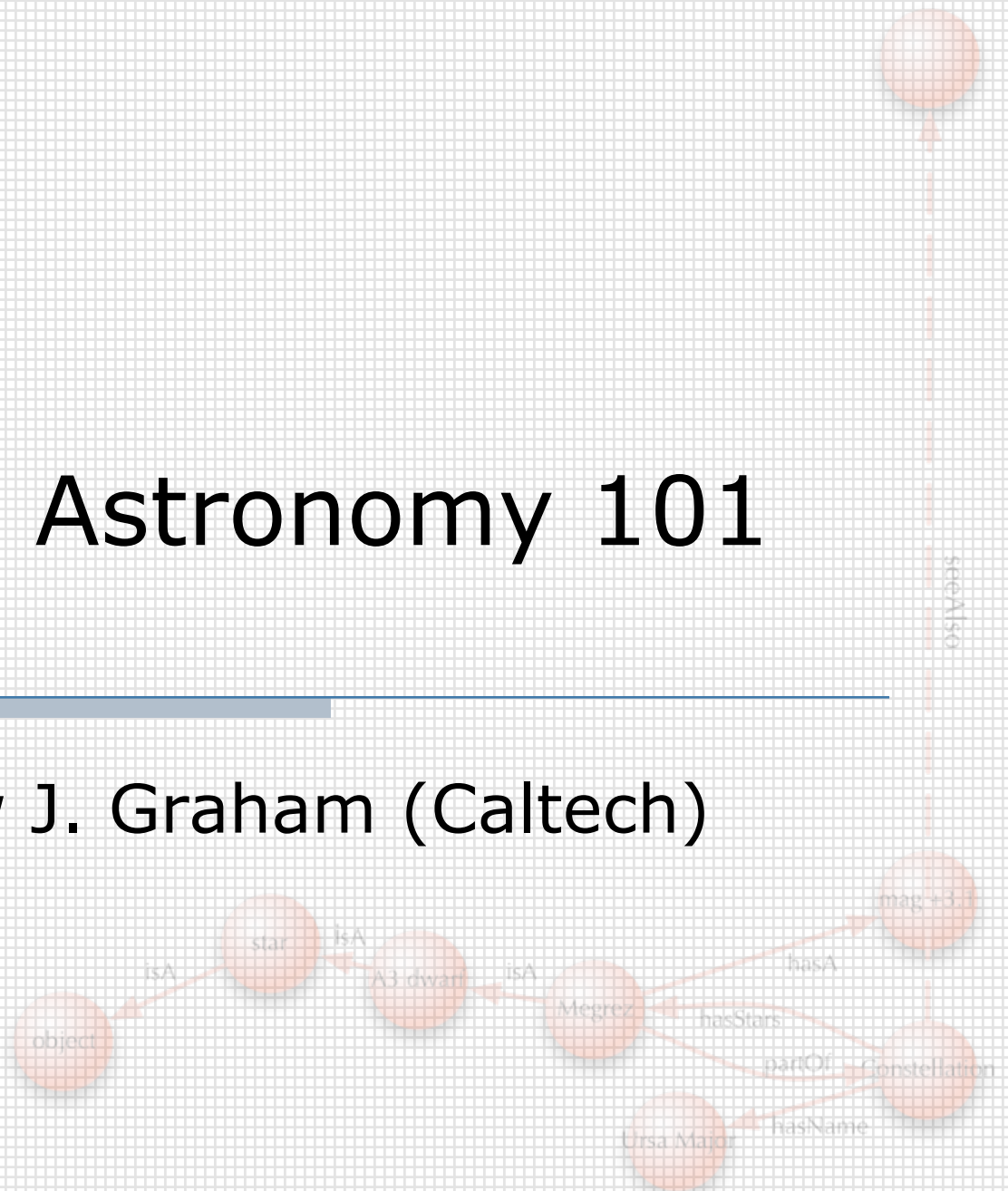


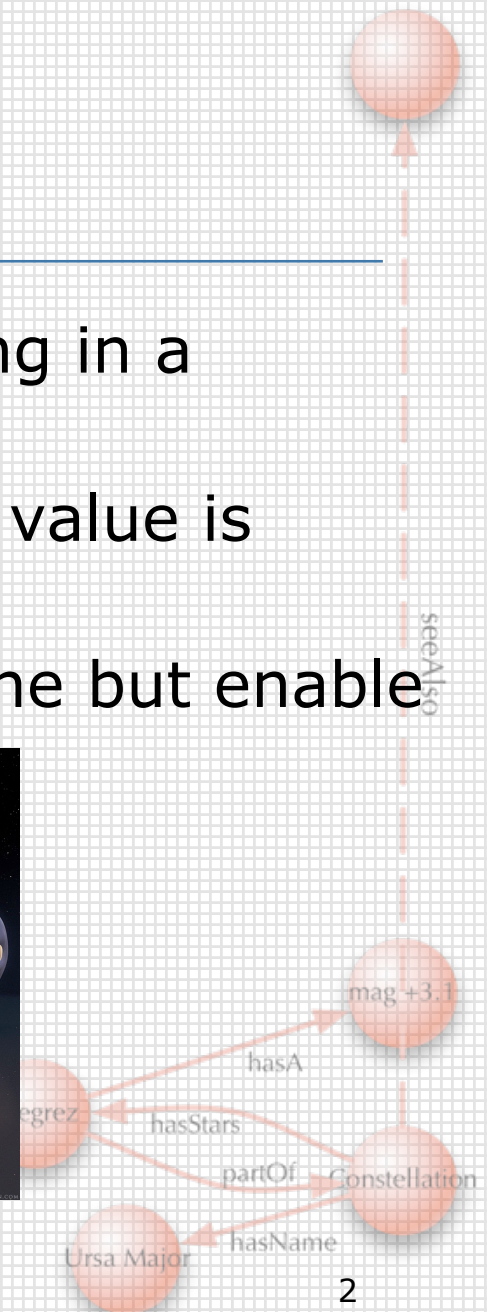
Semantic Astronomy 101

By Matthew J. Graham (Caltech)

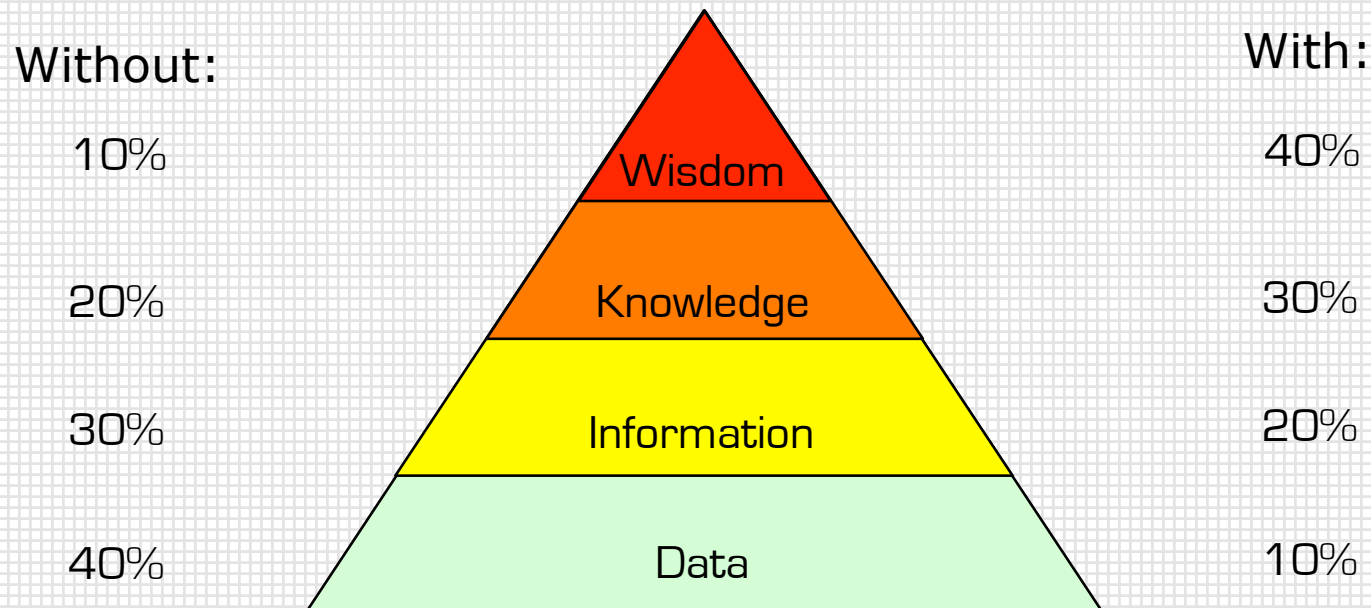


Semantics is...

- the *explicit* representation of meaning in a *machine-processible* form
- *ubiquitous* - all data above the pixel value is metadata
- a *community* effort – we do not define but enable definition
- *necessary*:



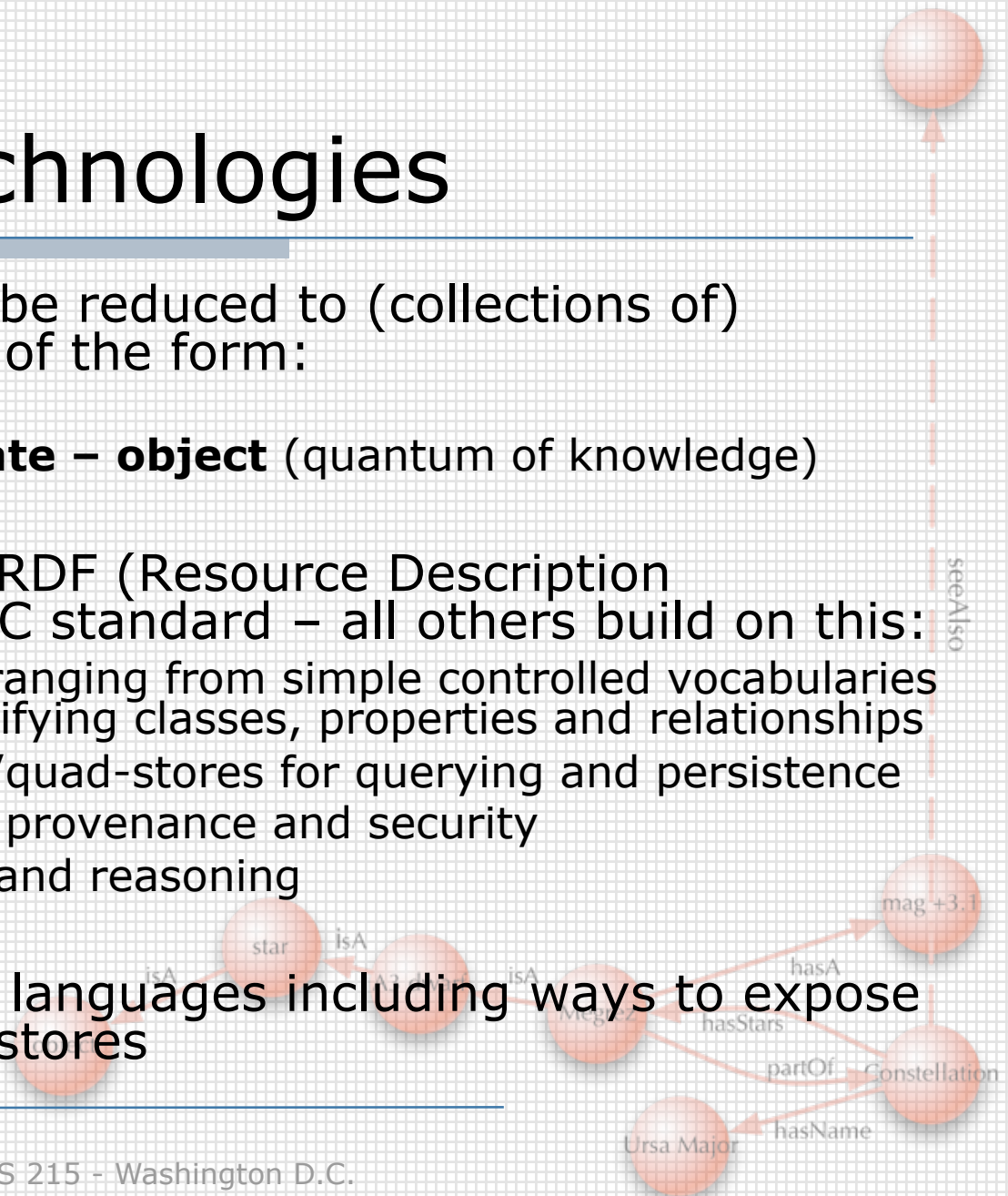
Benefits of semantics

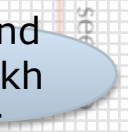


- Without technology, cognition mires in low-layer processes
- Semantic technology upgrades allocation of cognitive energy
- Semantics is essential in the data-intensive era

Semantic technologies

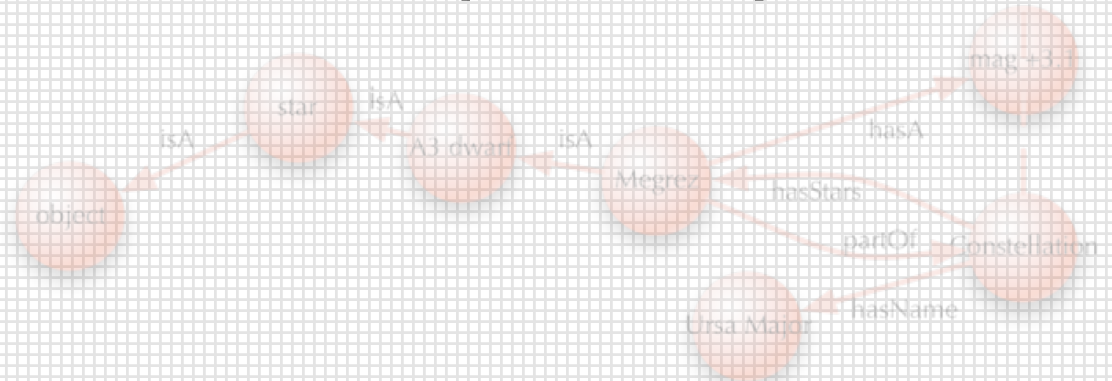
- All (meta)data can be reduced to (collections of) statements of fact of the form:
 - **subject – predicate – object** (quantum of knowledge)
- Core technology is RDF (Resource Description Framework), a W3C standard – all others build on this:
 - Concept schemes ranging from simple controlled vocabularies to ontologies specifying classes, properties and relationships
 - SPARQL and triple/quad-stores for querying and persistence
 - Proof and trust for provenance and security
 - Rules, inferencing and reasoning
- Supported in many languages including ways to expose existing relational stores





Where are we in 2010?

By Matthew J. Graham (Caltech)



Year 3 of the Glorious Revolution

- UK: 1690
 - Battle of the Boyne
 - Calcutta founded
 - First full peal rung
- USA: 1778
 - Valley Forge
 - Franco-American alliance
 - First US budget
- France: 1794/5 (FRC)
 - Thermidorian Reaction
 - Adoption of the metre
- Russia: 1919/20
 - Great Siberian Ice March
 - Recognition of Baltic states
 - GOELRO plan



Year 3 of the Semantic Revolution

2008: Practical Semantic Astronomy

Caltech, USA, 2008 February 18-21
http://www.cse.cmu.edu/astroinformatics

2009: PRACTICAL SEMANTIC ASTRONOMY 2009

GLASGOW, UK, 2-5 MARCH

The second Practical Semantic Astronomy Workshop will bring together experts from a broad range of disciplines using semantic technologies, alongside practitioners experimenting with these techniques, to address current problems in astroinformatics.

Topics:

- Astronomy and solar ontologies;
- Knowledgebases;
- Metadata for astronomical databases;
- Semantic integration;
- Semantic queries and data mining;
- Semantic technologies.

Call for papers deadline: 17 January 2009
Early registration deadline: 31 January 2009

2010: AstroInformatics2010

Caltech, USA, 2010 June 16-19, 2010

Defining an emerging discipline for the data-rich, computationally enabled astronomy in the 21st century

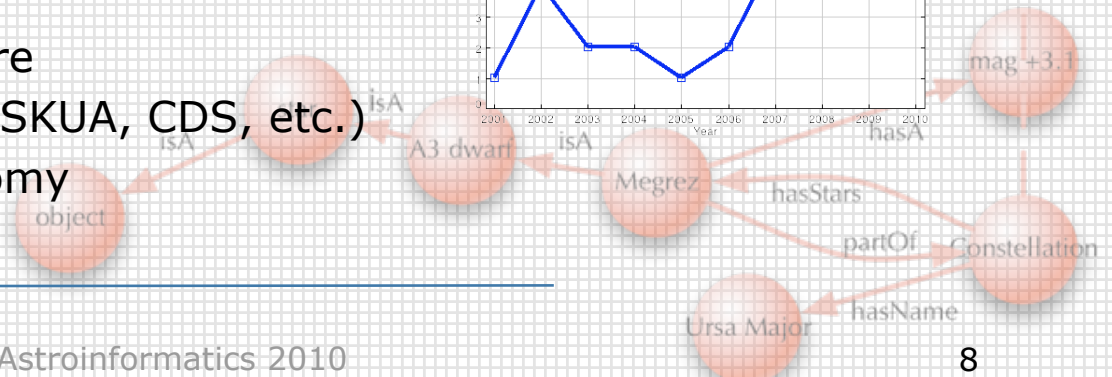
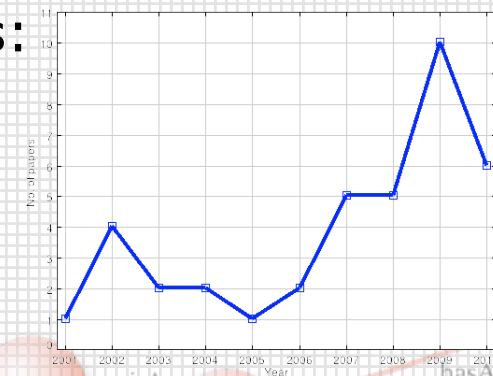
Topics include:

- Virtual observatories: the first decade and beyond
- Semantic technologies in astronomy and astrophysics
- Tools in computing, Cloud, IaaS, SaaS, and PaaS
- Knowledge discovery and extraction
- Data integration, federation, and virtualization
- Visualization of highly complex data sets
- Real-time analysis of astronomical data streams
- Semantic integration of astronomical data and data science technologies
- Collaborative environments and tools for astronomical data science
- Astronomical data science: the future of astronomy
- Semantic technologies in astronomy and astrophysics
- Tools in computing, Cloud, IaaS, SaaS, and PaaS
- Knowledge discovery and extraction
- Data integration, federation, and virtualization
- Visualization of highly complex data sets
- Real-time analysis of astronomical data streams
- Semantic integration of astronomical data and data science technologies
- Collaborative environments and tools for astronomical data science
- Astronomical data science: the future of astronomy

www.astroinformatics2010.org

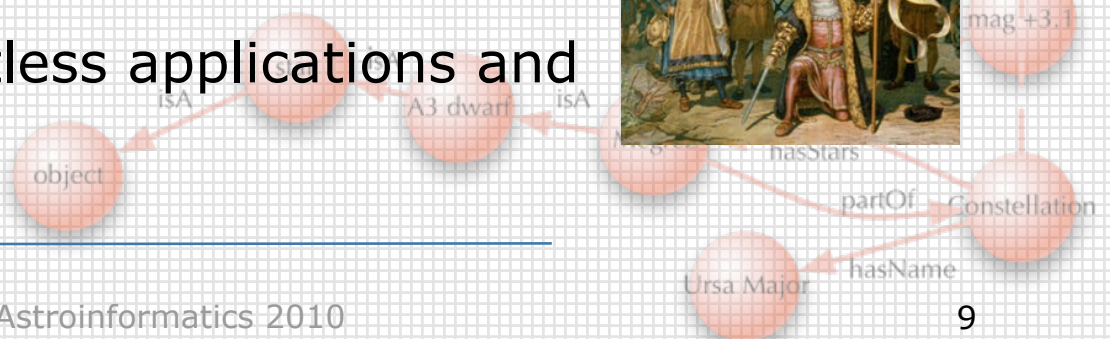
- IVOA:
 - Vocabularies using SKOS
 - Object type ontology
- Grass-roots efforts:
 - Astronomical literature
 - Annotation services (SKUA, CDS, etc.)
 - Time-domain astronomy
 - Data centers?

Publications:



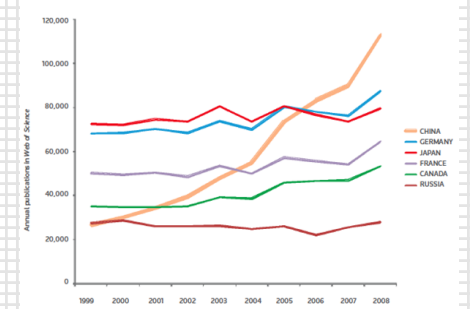
A history of semantic trailblazing

- Bioinformatics was born in 1977 with the sequencing of the bacteriophage Φ -X174
- Modern use of term “ontology” stems from 1991
- Widespread use in biomedicine by 2000
- Open Biomedical Ontologies (OBO) Consortium is now the umbrella organization coordinating semantic efforts in over 60 distinct domains of biology
 - A plethora of ontologies: upper, domain, formal, informal, reference, application
- Over 230 books on Amazon about ontologies in the biological sciences and numerous journals
- Underpinning countless applications and data mining efforts



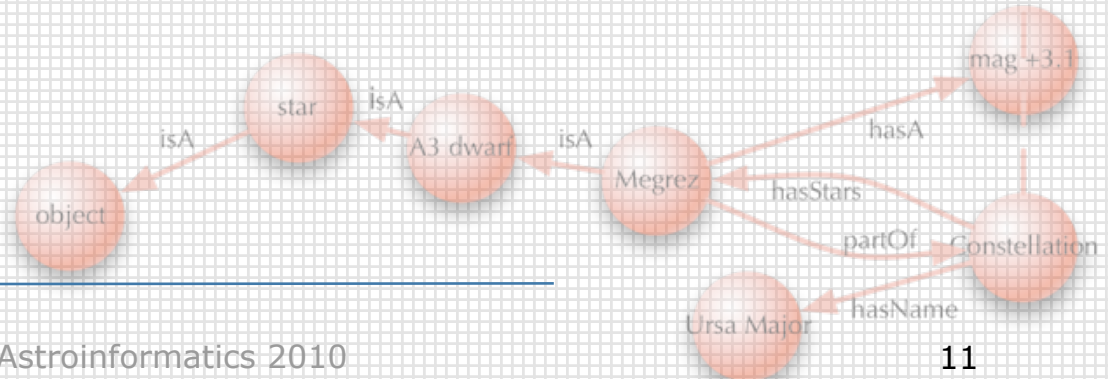
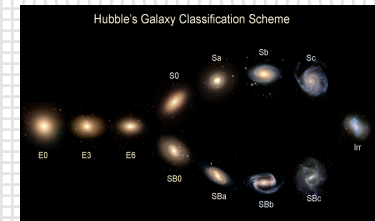
Why isn't astronomy the China of semantic sciences?

- Is it a lack of general awareness?
 - We've never heard of semantics
- Is it purely sociological/cultural?
 - Those techniques won't work here, our (meta)data is different
- Is it a lack of vision?
 - Why do we need a new approach now?
- Is it a lack of trust?
 - Computers cannot understand data
- Is it too trendy/cool?
 - Real science never uses exciting technology



The Linnaean problem

- Linnaeus' original system in 1735 already had 6 levels of hierarchy: 3 kingdoms, 35 classes, orders, genera, species, subspecies.
- Is astronomical knowledge still too coarse-grained to warrant the depth of modelling that an ontology can provide?
- Niche areas of taxonomic astronomy:
 - Solar system
 - “The Eurybates family is a compact core inside the Menelaus clan, located in the L_4 swarm of Jupiter Trojans.”
-arXiv:1004.4180
 - Exoplanets
 - Supernovae
 - Astrobiology



Data mining with nous

- Clustering
- Semantic similarity
- Detecting rare events via reasoning
- Fuzziness
- Association rules
- Network characterization

