

# The Astronomical eBrain

E. Shaya (U. of Maryland)



# Digital Library $\leq$ IVO

- VxO are in the class of e-Library systems
  - Modeled on Libraries, but on the web
  - Find Library/Repository/Archive
    - Business Registries/Yellow Pages  $\Rightarrow$  UDDI registries  $\Rightarrow$  VORegistry
  - Find Book/Dataset
    - Card catalogs  $\Rightarrow$  online card catalogs  $\Rightarrow$  Search the site pages (for tables or publication)
    - Dewey Decimal or LC Classification  $\Rightarrow$  dataset ID
  - Find Chapter/Column or Row
    - Book index  $\Rightarrow$  online linked index (DocBooks)  $\Rightarrow$  SQL/VOQL (to correct rows, or image), articles have index in left frame.
  - Services are new concept leveraging off of automation of above.

# Use Cases

- Column name="BHia" means Apparent brightness within the Holmberg isophotal radius, corrected for inclination, and absorption according to Hieles 21 cm data.
- Column name="Bia" – same thing except within isophotal radius of 26 mag/sq arcsec.
- Column name="Parallax" – but units are 100 mas per century.
- Theta – angular distance from Virgo Cluster
- pm means "proper motion" in the N30 system


# Use Case

- Look for “opacity of helium in the near-infrared”.
- Query: infrared “opacity of helium”
- Response: “Cool, helium-rich atmosphere white dwarfs are characterized by high atmospheric pressures resulting from the low opacity of helium” \* 1000
- Semantic Search:  
SELECT em:opacity  
em:opacity phys:ofPhysicalThing chem:helium;  
em:inBand em:nearInfrared.  
OR  
Chem:helium has:Quantity em:opacity
- We would also need to be able to do constraint on wavelength axis of an opacity spectrum.
- Note: Need to ensure that “low opacity of helium” above does not result in instance of opacityOfHelium.

# The Mind $\leq$ Semantic Web [What is a computer for?]

- Semantics ultimately will have machines read, talk, and do logical things (imitate thinking)
- Word Meaning  $\Rightarrow$  Ontology (Class, properties)
- Sentence  $\Rightarrow$  Triplet (Class, Property, Class) [imitate language]
- Questions  $\Rightarrow$  SparQL
  - What/Who?  $\Rightarrow$  (?s, p1, v1)
  - What does S1 p1?  $\Rightarrow$  (s1, p1, ?v)
  - What is the relationship between ..?  $\Rightarrow$  (s1, ?p, v1) [To form a query]
  - Why?  $\Rightarrow$  (statement1, "because", ?statement2)
  - How?  $\Rightarrow$  (statement1, "using", ?v)
- One is unaware of compartments within the brain  $\Rightarrow$  Semantic Web user does not specify data location.
- The mind remembers its sources and can ignore info from certain sources; likewise SW. If need be, users could exempt certain data sources.

# Semantic Knowledge Pools

- Images and Spectra are the skin of the orange, covers everything, but needs processing to digest. Tables are the inner rind, can digest but always insufficient. The meat is the knowledge in journals. Need to NLP journals into OWL.
  - Most ADS and many Google searches end with a bunch of journal articles to read through.
  - NLP to OWL of journals will mark new information age. Present search is nice, like a sports car. But SW search is like a fighter jet; it takes you to the answer of your question.
  - Storage of instances could be a tree following the ontology. Searches then traverse a few branches.
- 

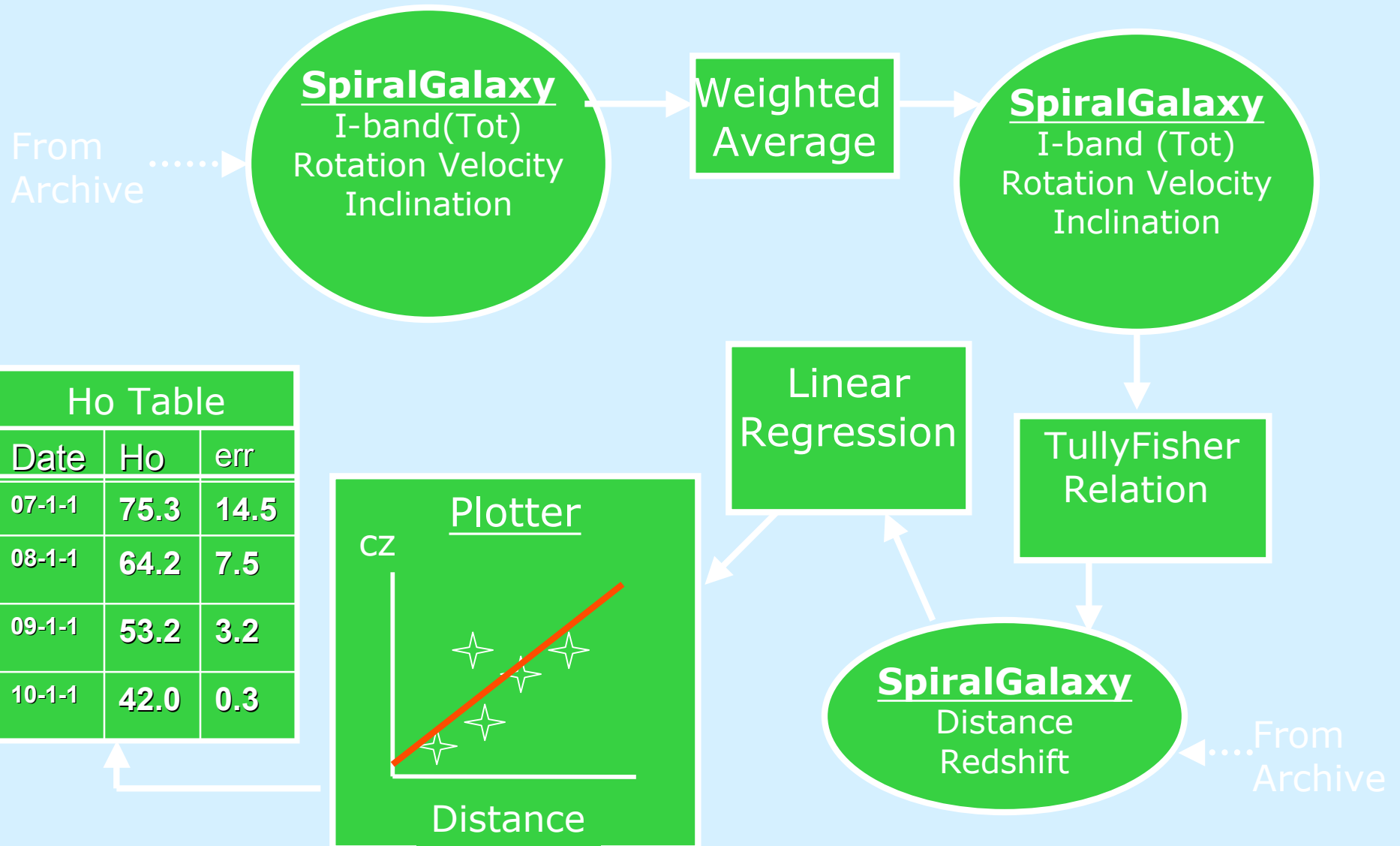
# Binding Terms to Code

- Machine understanding of terms
  - Machine does something special with each term
  - OWL allows Things to have properties with constraints.
    - Handle “star” differently from “galaxy” from “hasLuminosity”
    - Accept “RR\_Lyrae” as a “star” not “galaxy”
  - Bind to each term, inheritance just speeds up task:
    - Coordinates causes graphical maps are called
    - 3d\_coordinates causes 3 axis maps are called
    - EquatorialCoordinates have precession methods.
    - Units calls toSI and scaling methods
    - SI\_unites just calls scaling methods
    - Photometric Color calls subtraction methods
    - Quantities call error bar mathematics

# Functions as WebServices

- (Class1, Relation1, {Class2, Class3, Class4}) (Can be user specified)
  - Galaxies have Tully-Fisher (Spirals) or TRGB or Faber-Jackson Relationships (Ellipticals) . [statistical relations]
  - Luminosity is Surface Area \* Flux [exact relations]
  - Luminosity is proportional to  $M^{3.5}$ . [statistical relations]
  - The “relates” property tells which WebService binds to which Class.
- Merger of Analysis Toolkit plus Data Hunter Web Service
- The classes now form a network of related nodes and paths to any point can be deduced by network analysis. Except traversals usually require more than one node. (PLR: Period + Luminosity traverses to Distance)

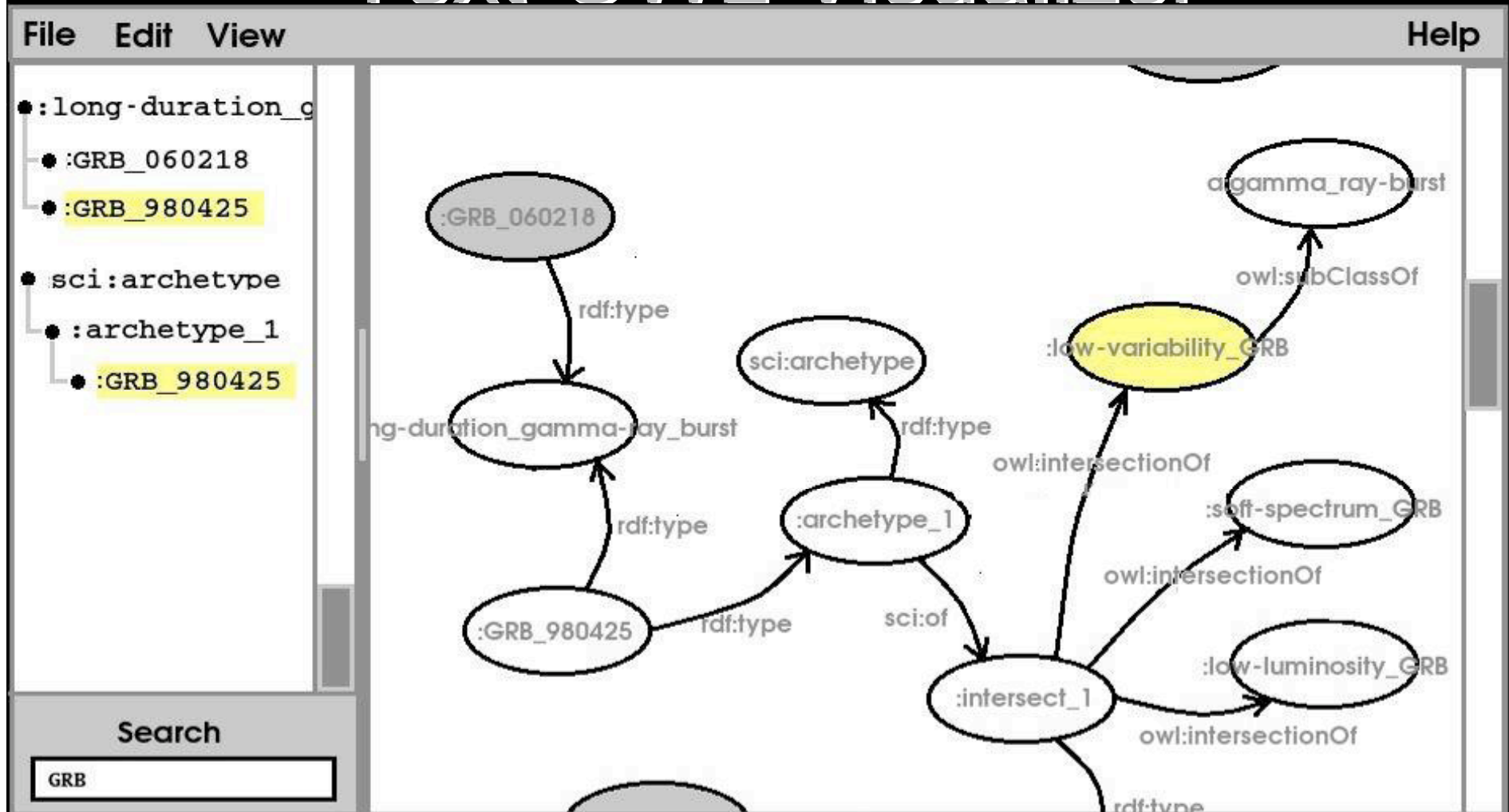
# H<sub>0</sub> Determination on OWLViper Canvas



# NLP of Journals

- In contrast, a semantics-enabled *knowledge base* holds networks of interrelated statements. These can provide deeper meaning behind the numbers by incorporating statements made from a rich domain terminology.
- As the knowledge base grows, reasoners (Racer, Pellet , etc.) can check for consistency and report on either all new data or new inconsistencies.
- benefits from the existence of rich astronomical thesauri presently being standardized by the IVO

# Text-OWL Visualizer



Indeed, GRB 980425 is usually taken as the archetypal low-luminosity, low-variability, soft spectrum GRB.

# Merger of Facts

- How to deal with repetitive equivalent facts?
- Waste of cpu+bandwidth to search and retrieve same fact held in multiple repositories. Waste to search multiple repositories with very similar info.
  - Possible sol'ns: shadow out regions of repositories, or aggregate info elsewhere.
- How to deal with slightly different facts?
  - Fact subfact of another – Star L < 1e3Msuns, Star L < 1e2 Msuns. – needs cleanup
  - Agreement but different, ie within errorbars of each other – needs user selected handling
- Handling disagreement
  - Star is G-star vs Star is K-star. Galaxy is Sab vs Galaxy is Sa.
    - OWL fatal error if classes are disjoint.

# Entity Pools

- Query: “Class P1; P2; P3.”
- Drag Instances of P1 from DB1 to DB2.  
Drag P1; P2 to DB3. Return
- Eventually we get tired of waiting.
- Galaxies at NED. Stars at SIMBAD (not stars in galaxies?). Atomic and Molecular Physics at ?

# The Web vs The Desktop

- How much should one rely on one's mind vs how consulting the references?
  - In this Corner we have The Web, weighing in at ?
    - High latency and low bandwidth.
    - Capable of storing exabytes of data.
    - Usually up to date.
  - And in this Corner we have The Desktop
    - Capable of storing terabytes of data
    - Low latency and high bandwidth
  - Need to create system that intelligently uses both. Desktop becomes another knowledge pool.
  - Is the timeliness of the data critical?
  - Might there be more data or instances or sources of data since last time?
  - Do we re-ask or do we supplement.
    - I asked this last Wednesday. Anything new here?

# Conclusion

- Astronomy eBrain is a kind of VOSpace where semantic analysis plus functional analysis goes on together. The library structures are invisible to the user.
- Enable web services that reach scientists final goal while reducing human interaction; will evolve into autonomy.
- We need discuss how to take strides rather than inch along.