



VIRTUAL ASTRONOMICAL OBSERVATORY

Smart Applications in Biology

Matthew J. Graham, Caltech

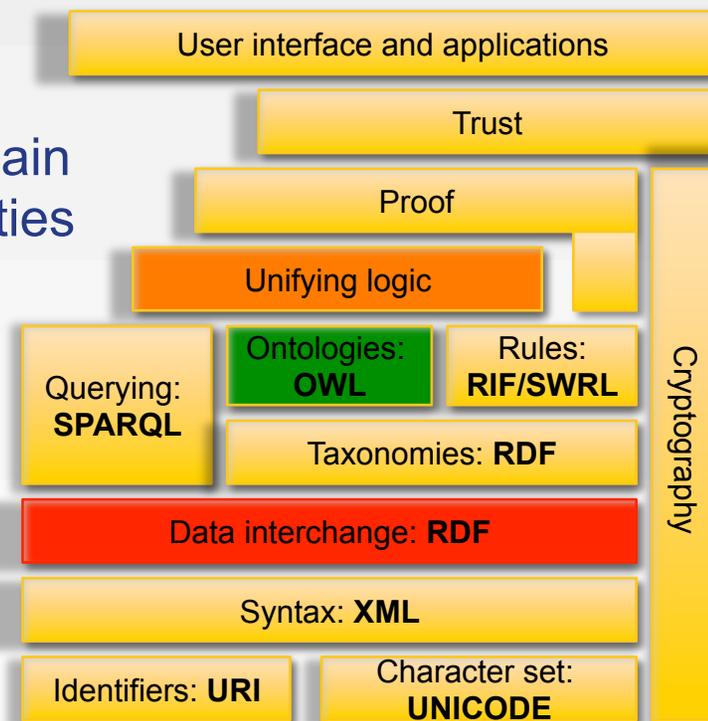


The VAO is operated by the VAO, LLC.



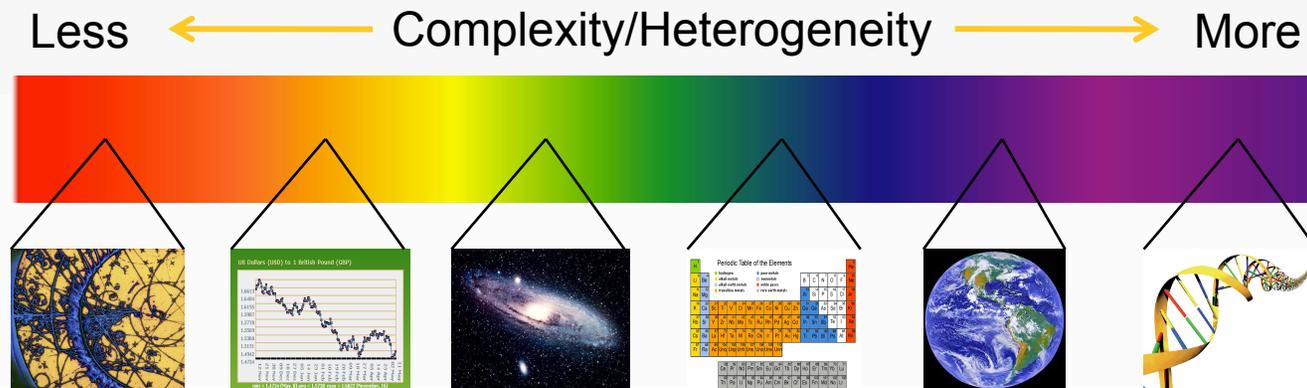
What is a smart application?

- One built around technologies that *understand* data and *know* or can *infer* what to do with it
- What makes things smart?
 - **RDF**: all data can be represented as subject – predicate – object
 - **Ontology**: a conceptual model of domain knowledge in terms of classes, properties and relationships
 - **Description logic**: the backbone for inferencing and checking instances, relations, subsumption and concept consistency



Why is smartness prevalent in biology?

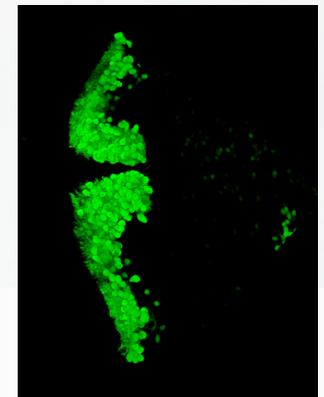
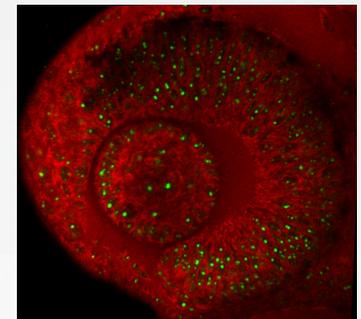
- X-informatics is the discipline of organizing, accessing, mining and analyzing information describing complex systems in (x = bio-, geo-, chemo-, astro-, econo-, ...)
- Bioinformatics was born in 1977 with the sequencing of the bacteriophage Φ -X174
- Developments in genomic and information technology have produced a huge amount of complex and disparate *information*
- Smartness introduced via semantic technologies to address this





The Zebrafish FlipTrap data repository

- A systems-based approach for analysis of gene function in developing vertebrate embryos in real time and space
- The FlipTrap screen is a gene trap that fuses the Citrine fluorescent protein to the trapped protein to generate a fully functional tagged version
- Expression patterns of the marked gene during development can then be imaged, etc.
- The data repository holds images, metadata, sequence data and annotations
- It makes extensive use of the Zebrafish anatomical ontology (2400 classes, 8 properties, 11038 entity annotation axioms) and the Gene ontology (30393 terms - 99.2% with definitions incl. 18939 biological process, 2735 cellular component and 8719 molecular function)





Example: smart data entry

Center of Excellence in C

Summary
Image data
Molecular data
Functional data

Screen 1

Generation:

Day screened:

Stage:

Site of expression:

Choose one:

-
-
-

Subcellular expression:

-
-
-
-
-
-
-
-
-
-

Additional comments:

No. of filesets to upload:

Only those anatomical structures defined by the ontology to be present at the selected "stage" are available for selection in the autocomplete drop-down.



Example: smart querying

ZEBRAFISH FLIPTRAP DATABASE
Center of Excellence in Genomic Science

SEARCH:

BROWSE BY

- Anatomical Location
- Stage
- Gene Name
- Allele
- Multiple Parameters

DOCUMENTS

- Contact Us

RESOURCES

- Additional Genomic Resources

Multiple Parameter Search

Expand one or more sections to enter search criteria and then click the Search button below.

Stage

<input type="checkbox"/> Zygote: 1-cell	<input type="checkbox"/> Gastrula	<input type="checkbox"/> Hatching
<input type="checkbox"/> Cleavage	<input type="checkbox"/> 50%-epiboly	<input type="checkbox"/> Long-pec
<input type="checkbox"/> 2-cell	<input type="checkbox"/> Germ-ring	<input type="checkbox"/> Pec-fin
<input type="checkbox"/> 4-cell	<input type="checkbox"/> Shield	<input type="checkbox"/> Larval
<input type="checkbox"/> 8-cell	<input type="checkbox"/> 75%-epiboly	<input type="checkbox"/> Protruding-mouth
<input type="checkbox"/> 16-cell	<input type="checkbox"/> 90%-epiboly	<input type="checkbox"/> Day 4
<input type="checkbox"/> 32-cell	<input type="checkbox"/> Bud	<input type="checkbox"/> Day 5
<input type="checkbox"/> 64-cell	<input type="checkbox"/> Segmentation	<input type="checkbox"/> Day 6
<input type="checkbox"/> Blastula	<input type="checkbox"/> 1-4 somites	<input type="checkbox"/> Days 7-13
<input type="checkbox"/> 128-cell	<input type="checkbox"/> 5-9 somites	<input type="checkbox"/> Days 14-20
<input type="checkbox"/> 256-cell	<input type="checkbox"/> 10-13 somites	<input type="checkbox"/> Days 21-29
<input type="checkbox"/> 512-cell	<input type="checkbox"/> 14-19 somites	<input type="checkbox"/> Juvenile
<input type="checkbox"/> 1k-cell	<input type="checkbox"/> 20-25 somites	<input type="checkbox"/> Days 30-44
<input type="checkbox"/> High	<input type="checkbox"/> 26+ somites	<input type="checkbox"/> Days 45-89
<input type="checkbox"/> Oblong	<input type="checkbox"/> Pharyngula	<input type="checkbox"/> Adult
<input type="checkbox"/> Sphere	<input type="checkbox"/> Prim-5	<input type="checkbox"/> Unknown
<input type="checkbox"/> Dome	<input type="checkbox"/> Prim-15	
<input type="checkbox"/> 30%-epiboly	<input type="checkbox"/> Prim-25	
	<input type="checkbox"/> High-pec	

Anatomical Location

<input type="checkbox"/> Zebrafish Anatomical Entity	<input type="checkbox"/> Anatomical Structure	<input type="checkbox"/> Portion Of Organism Substance
<input type="checkbox"/> Anatomical Line	<input type="checkbox"/> Acellular Anatomical Structure <input type="checkbox"/>	<input type="checkbox"/> Blood
<input type="checkbox"/> Groove <input type="checkbox"/>	<input type="checkbox"/> Anatomical Group <input type="checkbox"/>	<input type="checkbox"/> Cerebral Spinal Fluid
<input type="checkbox"/> Anatomical Space	<input type="checkbox"/> Cell <input type="checkbox"/>	<input type="checkbox"/> Dentine
<input type="checkbox"/> Bile Canaliculus	<input type="checkbox"/> Compound Organ <input type="checkbox"/>	<input type="checkbox"/> Enameloid
<input type="checkbox"/> Choroidal Fissure	<input type="checkbox"/> Embryonic Structure <input type="checkbox"/>	<input type="checkbox"/> Otolith <input type="checkbox"/>
<input type="checkbox"/> Coelom	<input type="checkbox"/> Extraembryonic Structure <input type="checkbox"/>	<input type="checkbox"/> Synovial Fluid
<input type="checkbox"/> Opercular Cavity	<input type="checkbox"/> Multi-tissue Structure <input type="checkbox"/>	<input type="checkbox"/> Unspecified
<input type="checkbox"/> Pericardial Cavity	<input type="checkbox"/> Organism Subdivision <input type="checkbox"/>	
<input type="checkbox"/> Pleuroperitoneal Cavity	<input type="checkbox"/> Portion Of Tissue <input type="checkbox"/>	
<input type="checkbox"/> Pupil	<input type="checkbox"/> Whole Organism <input type="checkbox"/>	

Terms and hierarchies generated dynamically from the ontology

Anatomical structures not present at the selected stage will be grayed out



Example: smart results

Automatic links to literature about this gene

Full resolution images via appropriate viewer available at a click

Sequence information with exons highlighted and location of citrine marker indicated

Gene Expression Report

Allele	ft122a	
ZFIN	ZDB-GENE-980526-221	
NCBI		Ensembl
Gene Name	desm	desm
Aliases	des cb290 fb59a12 MGC109859 wu:fb59a12 zgc:109859 desmin	cb290 des fb59a12 wu:fb59a12 zgc:109859
Gene Description	desmin	desmin [Source:RefSeq peptide;Acc:NP_571038]
Gene Ontology	show	show
NCBI Nucleotide Hit	NM_130963 show details	
Ensembl Genomic Map		
Comments	<pre> hide CATTACACAGCGTAC ACCGCCGCACCTTGG GACTGACCTCCAGAGT TCGGAGGTGGCTCGGT AGGACTTCCTCAACAC AGGTGGCTTCCTCGA TTGCAGAGCTGTACGA AGATCGAGAGGGACAA AAGCTGAAAACAACCT TCGAGGCTCTTCACGA AGGAGAGTCAGCTGCA ACGAGGCTATCGCTGC TGAACAAGANTAACGA GCGAGATTGACTCTCT AGGCCGGTGGTTATCA TCCGGAGTACCAGGA GAGAGGAGAGCAGGAT ACCACCAGCAGCAGCA CCGATGGCGAGTCTGT CAGAATGCCTTGCATT TGTGGATGACGCATGG CGTTCAGAAAGTGGCC </pre>	
mRNA sequence		



Other smarts

- Suggestions
 - Search for related data products based on semantic similarity
- Environments
 - Virtual lab books linked to data and literature
 - Shared workflows with myExperiment.org
- Data mining
 - Incorporating domain knowledge into the discovery process





Smart applications in astronomy?

- Linked data
 - ADS, CDS, NED
 - 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, exoplanets, supernovae?
- “The Eurybates family is a compact core inside the Menelaus clan, located in the L_4 swarm of Jupiter Trojans.”

– arXiv:1004.4180

