VIRTUAL ASTRONOMICAL OBSERVATORY

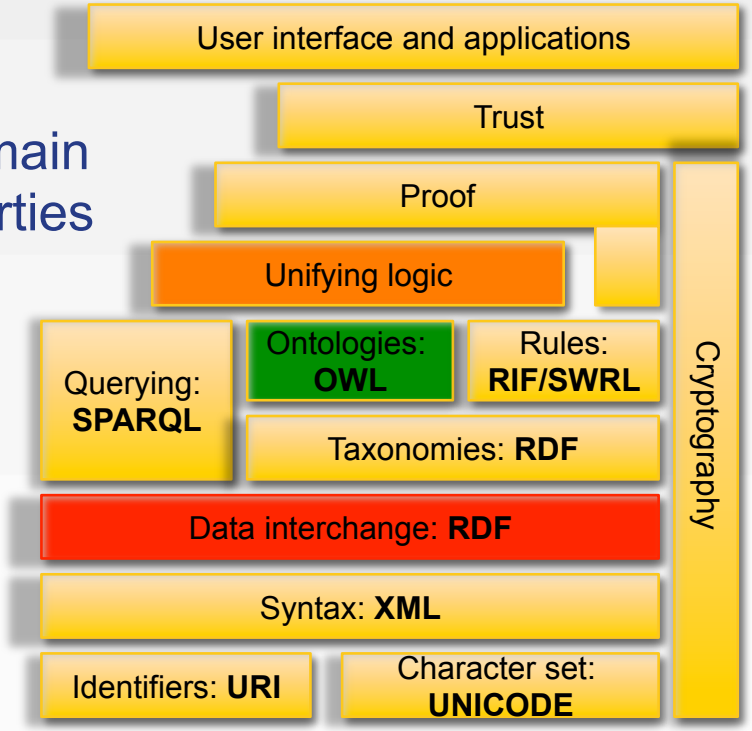
Smart Applications in Biology

Matthew J. Graham, Caltech

What is a smart application?

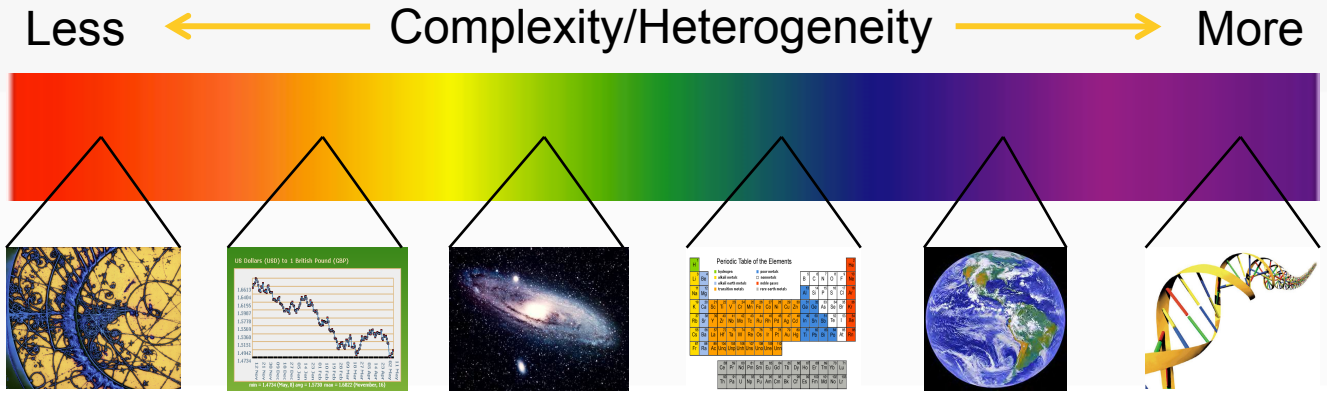
- One built around technologies that *understand* data and *know* or can *infer* what to do with it

- 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



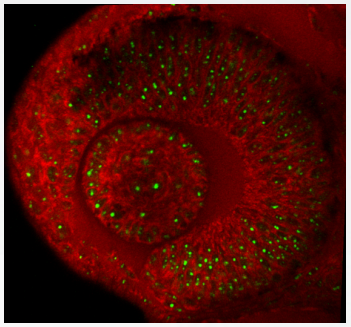
Why is smartness prevalent in biology?

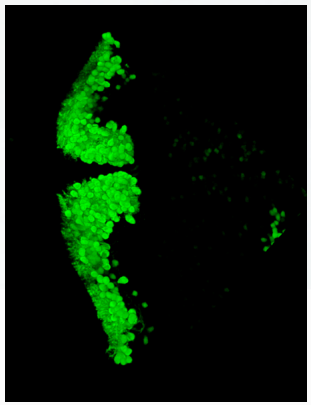
- X-informatics is the discipline of organizing, accessing, mining and analyzing information describing complex systems in (x = bio-, geo-, chemo-, astro-, econo-, …)
- Bioinformatics was born in 1977 with the sequencing of the bacteriophage Φ-X174
- Developments in genomic and information technology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



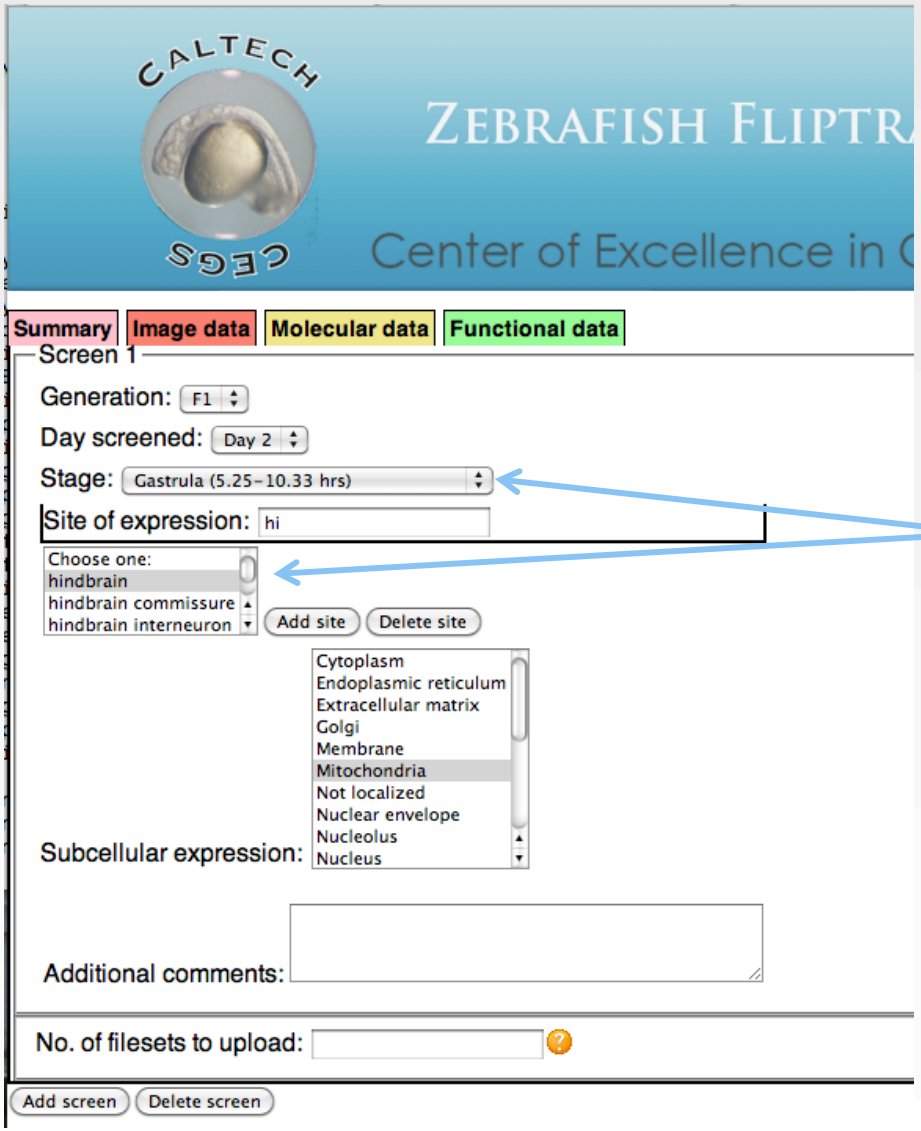
The Zebrafish FlipTrap data repository

- 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)



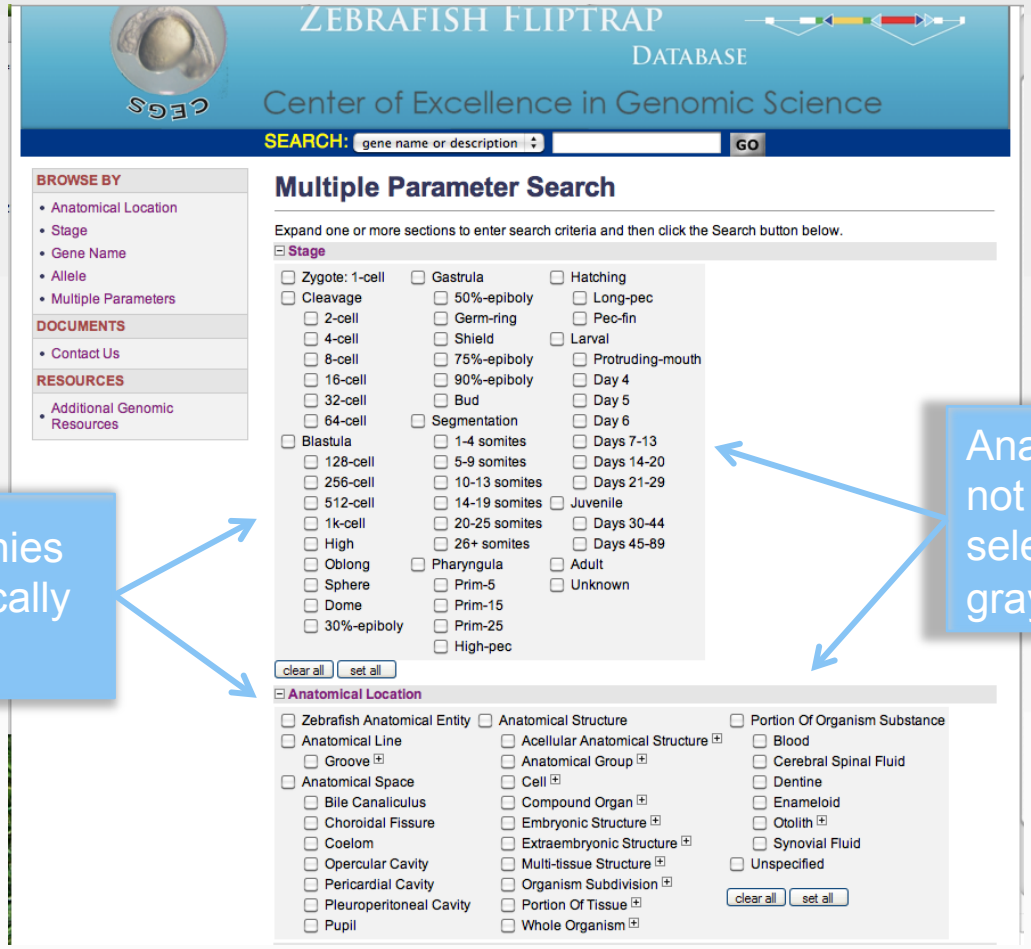


Example: smart data entry



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



Terms and hierarchies generated dynamically from the ontology

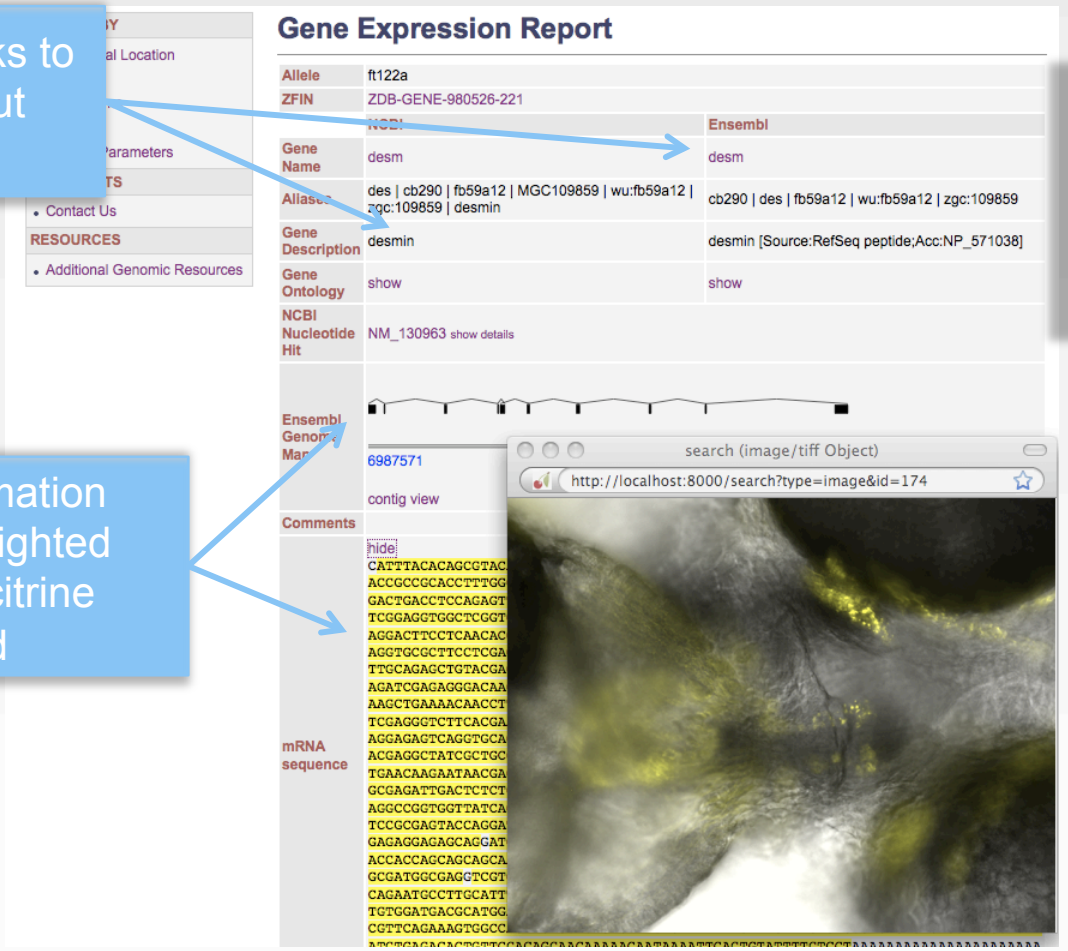
Anatomical structures not present at the selected stage will be grayed out

Example: smart results

Automatic links to literature about this gene

Sequence information with exons highlighted and location of citrine marker indicated

Full resolution images via appropriate viewer available at a click



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 taxonometric 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