Taking a view on bio-ontologies

Simon Jupp
Functional Genomics Production Team
ICBO, 2012
Graz, Austria
Who we are

- European Bioinformatics Institute one of world’s largest bio data and service providers

- Functional Genomics runs ArrayExpress, Gene Expression Atlas and Biosample Database
  - Adding value to data through Semantic annotation
  - Add value to data through visualisation tools

- Data ~30k experiments, 882k assays
  - 142K unique sample annotations
  - 120K unique assay annotations
Experimental Factor Ontology

- We consume parts of reference ontologies from domain
- Construct new classes and relations to answer our use cases
- Aim is reuse of existing resources, shared frameworks and mapping of equivalencies where they exist

Ontology Biomedical Investigations
- Disease Ontology
- Cell type /lines

Gene Ontology

Chemical Entities of Biological Interest (ChEBI)

http://www.ebi.ac.uk/efo
Experimental Factor Ontology – many views

Multiple Views

Ontology development

Database curators

Data submission

External ontologies

Analysis

protégé

Master headline
When user type “co” in search box they don’t want to see “continuant” in the drop down!
EFO in Applications

Sensible faceted browsing
Data analysis

• Ontology based annotations for sample clustering

User feedback

“Can you hide some of these cell type categories like “cell in vivo”, they are too generic to be useful.” – KUPKB user

Julie Klein, Simon Jupp, et al. The KUPKB: a novel Web application to access multiomics data on kidney disease
Ontology views

“I like the look of EFO, it has a lot of terms I’m interested in,
But other bits are not so interesting. Can I have a subset, plus these extra terms”
– some EFO consumer

“we really like the FMA, but it is too X for our needs, we really only need Y”. X is something like "too large" or "too detailed” –
http://sig.biostr.washington.edu/projects/ontviews/

A definition for a view…

“A collection of terms from one or more ontologies that are organised in a way that is useful for a particular community or application”
Types of views

- Modularisation (logical subsets)
  - OWL imports
- Query based
  - DL query / OPPL (get me all human subset)
  - SPARQL (construct queries)
    - vSparql (SIG UW)
- MIREOT (methodology)
- Annotation based
  - OntoDog (http://ontodog.hegroup.org)
  - OBO slims
EFO annotation based views

- Annotate hidden classes
  - EFO:organizational_class “true” e.g. “material entity” from BFO
- Assign classes to a subset/view
  - e.g. EFO:gwas_trait “true” e.g. for GWAS specific terms
- Root or branching classes
  - EFO:branch_class “true” e.g. organism
EFO view

- Hierarchical relationships
Looking for standards

• Bespoke code to extract views
  • Duplicated across applications
  • Need for a cleaner separation of knowledge representation from presentations
  • More explicit view semantics

• Everybody does it differently
  • No generic tooling
Exploring SKOS

- Simple Knowledge Organization System
  - Simple W3C vocabulary for thesauri, classification schemes, taxonomies, subject heading
- Using new OWL 2 features – punning
  - A way to index terms in our ontology
  - We can safely mix OWL 2 and SKOS
  - Meta-modeling with classes and individuals

- EFO Requirements – using SKOS to index EFO classes
  - Define a view ✓ (SKOS Concept Scheme)
  - View membership ✓ (SKOS in scheme property)
  - Branching classes ✓ (SKOS top concept)
  - Hierarchical relationships ✓ (SKOS broader / narrower properties)
Protégé 4 support

- SKOSEd plugin for Protégé 4.1
  - [http://code.google.com/p/skoseditor/](http://code.google.com/p/skoseditor/)
SKOSify EFO

owl:Class

skos:Concept

"Cardiac arrhythmia"

"cardiovascular disease"

skos:hasTopConcept

skos:inScheme

EFO:gwas_scheme

rdf:type

rdfs:subClassOf

skos:broader
NHGRI GWAS diagram generation
Concept Schemes

owl:Class

skos:Concept

EFO:EFO_0000764

"HIV"

EFO:EFO_0000767

"HIV prevention"

skos:ConceptScheme

Ex:MyGWASview

skos:topConceptOf

rdfs:label

skos:inScheme

skos:narrower

skos:inScheme

rdf:type

rdf:type

rdfs:label
Punning

http://www.ebi.ac.uk/efo/EFO_0004269
Viewgen framework (Under development)

- Java API for extracting ontology views
  - Generate OWL or SKOS exports
  - OWL DL, SPARQL or Annotation based
  - Protégé 4.1 plugin

- Future work
  - Web UI (term shopping cart)
  - Jquery widget library for embedding ontology views in web application
  - Other requirements?
SKOS may not be the answer

- Introduces lots of individuals (via punning)
- Conversion is lossy (semantics get a lot weaker)
  - Fine for visualization and supporting query expansion
  - Not so great for analysis
  - What do you do with anonymous class expressions?
- We want view specific labels, not supported by SKOS

\[\text{efo:EFO_0004160}\]

"preferred name"

"whole chromosome random sequencing"

"SRA name"

"WCS (Random sequencing of a whole chromosome or other replicon isolated from a genome)"
OWL annotations

• Is there enough interest to extend the OWL vocabulary?
  • OWL 2 rich annotations e.g. owl:deprecatedClass
  • owl:inSubset, owl:subsetRoot, owl:hierarchicalProperty

• Possible forum for discussion in the OWLEd community group
Summary

- The more data we annotate, the more ontology we have to deal with and expose through our user facing applications
  - Current ontologies are not fit for user consumption

- New communities are wanting slices of ontology
  - Metagenomics, Ensembl, ENA, GWAS …

- Current tools are not sufficient to manage this layer

- We need support for lighter-weight representations
  - SKOS provides a viable mechanism and has benefits in certain scenarios
Acknowledgements

- James Malone, Tony Burdett, Dani Welter, Helen Parkinson (EBI, Cambridge)
- Robert Stevens (University of Manchester)
- Andrew Gibson (Academic Medical Center, Amsterdam)

- Acknowledge funding from
  - EMBL (JM, TB, HP)
  - NHGRI GWAS (U41-HG006104)
  - NIH NCBO DBP (U54-HG004028)
  - EU-FP6 SeaLife (IST-2006-027269)