

Taking a view on bio-ontologies

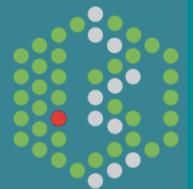
Simon Jupp

Functional Genomics Production Team

ICBO, 2012

Graz, Austria

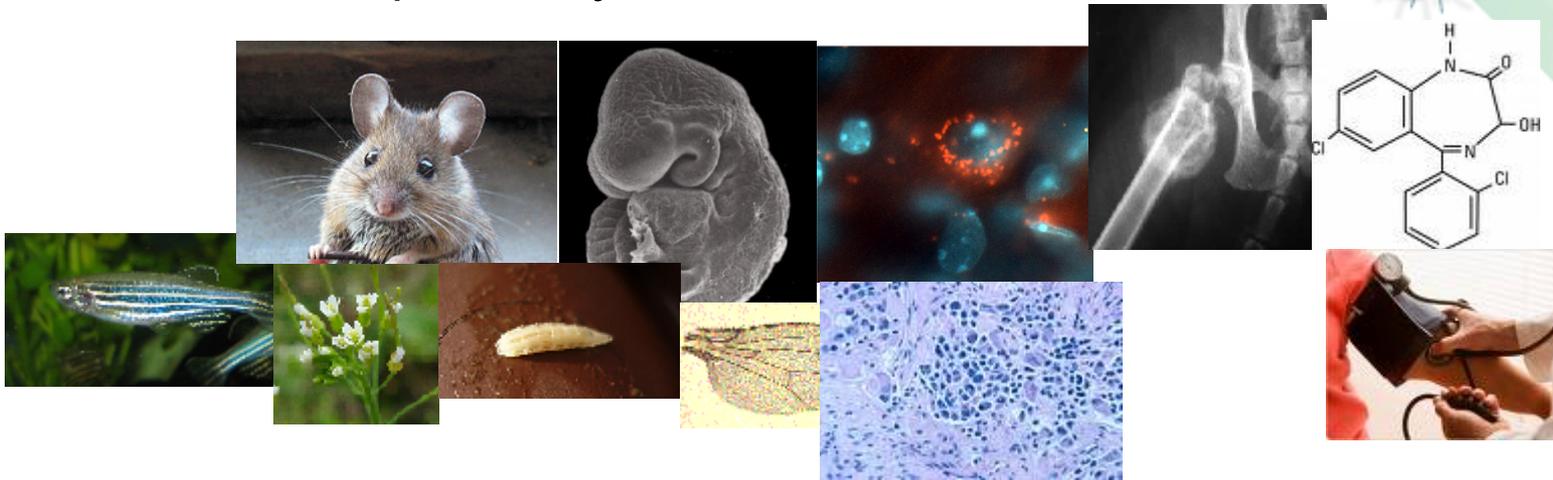
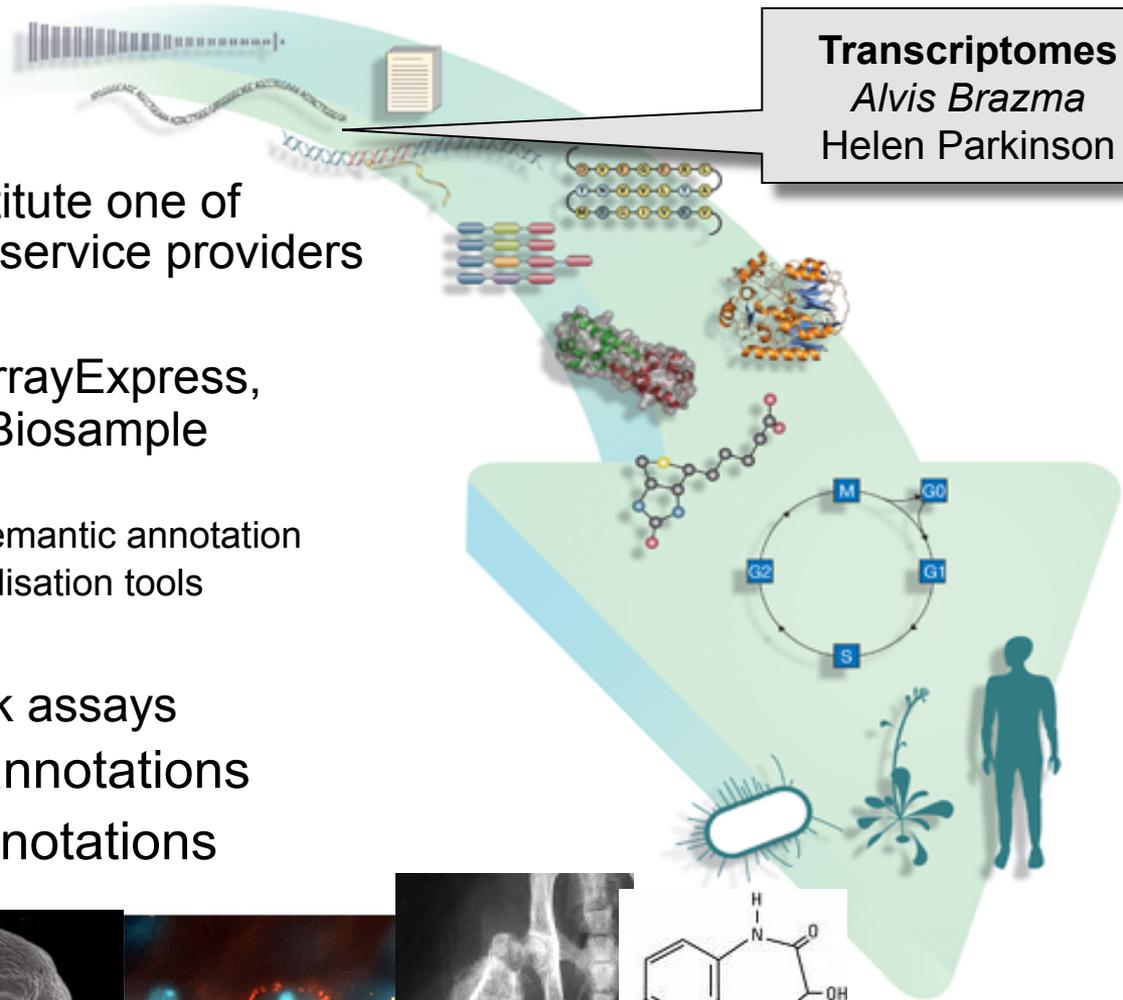
EMBL-EBI



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

Transcriptomes
Alvis Brazma
Helen Parkinson



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

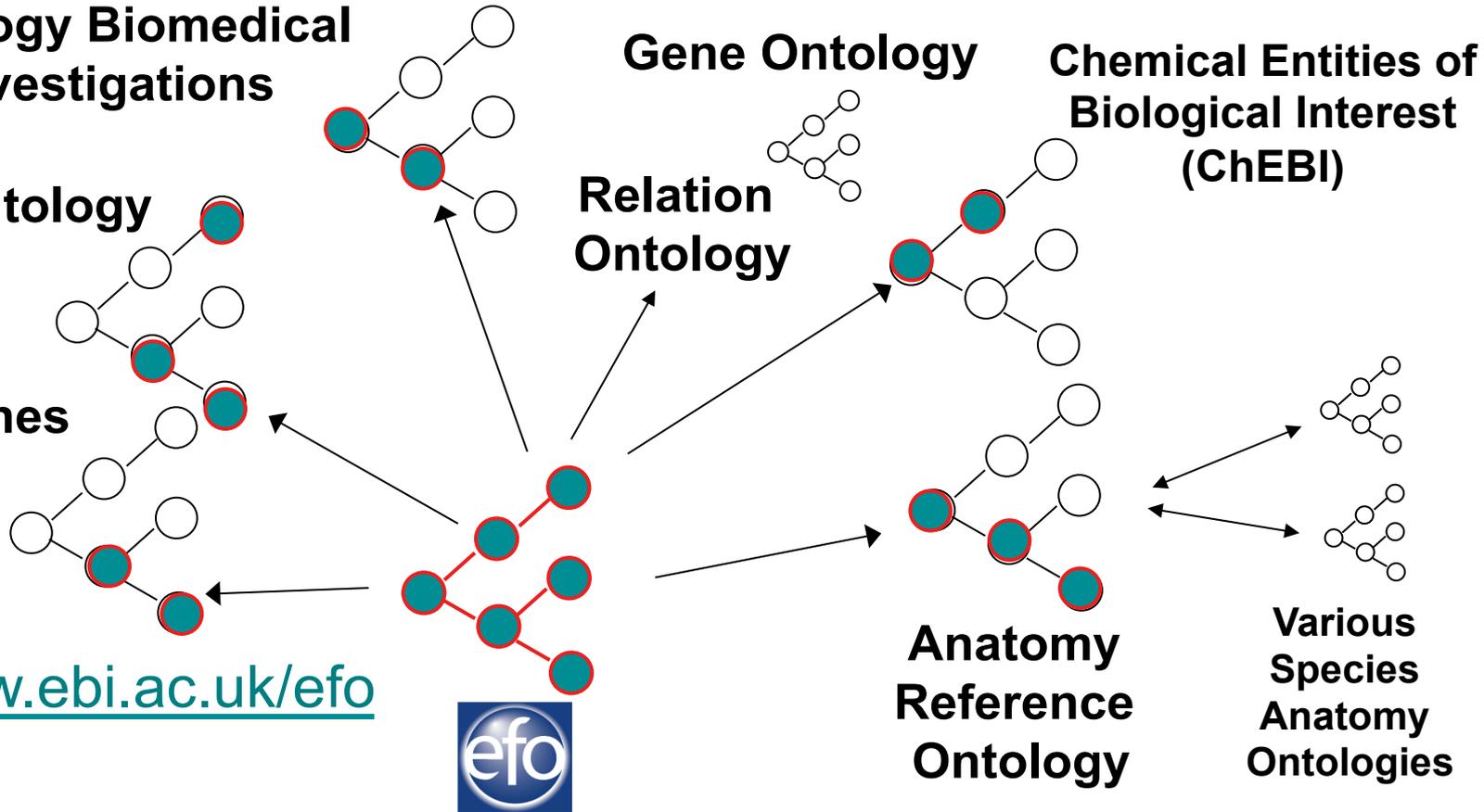
Relation Ontology

Chemical Entities of Biological Interest (ChEBI)

Anatomy Reference Ontology

Various Species Anatomy Ontologies

<http://www.ebi.ac.uk/efo>



Experimental Factor Ontology – many views

External ontologies



Ontology development



Multiple Views

Analysis

ATLAS

A screenshot of the ATLAS web interface. At the top, there are search filters for Genes (set to '(all genes)'), Organism (set to '(any)'), and Conditions. Below the filters is a search button labeled 'Search Atlas'. The main content area shows a table with columns for Gene, Organism, and Assay. Below the table is a 'GWAS Diagram Browser' showing a grid of colored bars representing GWAS results. The interface also includes a legend and a 'new' button.

Database curators



Data submission



EFO view in Applications

Experiment, citation, sample and factor annotations [clear]

leuk

leukaemia

- + leukemia
 - acute lymphoblastic leukemia
 - B-cell acute lymphoblastic leukemia
 - T-cell acute lymphoblastic leukemia
 - + acute myeloid leukemia
 - chronic lymphocytic leukemia

Experiment, citation, sample and factor annotations [clear] Filter

leukemia Any

Loaded in Gene Expression Atlas

Expand search using Experimental Factor Ontology

Submitter/reviewer login [ArrayExpress Browser Help](#)

Any array Detailed view

Any experiment type

ID	Title	Assays	Species	Date
+ E-TABM-763	MicroRNA profiling of peripheral blood mononuclear cells from CLL patients to i...	61	Homo sapiens	2009-08-14
+ E-TABM-762	MicroRNA profiling of peripheral blood mononuclear cells from CLL patients to i...	97	Homo sapiens	2009-08-14
+ E-TABM-726	Transcription profiling of liver from wild type, Rev knock out and Rev-erb-alpha...	6	Mus musculus	2009-09-03
+ E-TABM-696	Transcription profiling of human Chronic Lymphocytic Leukemia cells with muta...	24	Homo sapiens	2009-04-30
+ E-TABM-695	Transcription profiling of mouse LSK hematopoietic stem cells from wild type a...	12	Mus musculus	2009-04-28
+ E-TABM-694	Transcription profiling of mouse LT-HSC hematopoietic stem cells from wild typ...	5	Mus musculus	2009-04-28
+ E-TABM-670	Transcription profiling of mouse embryonic stem cell line CGR8 grown in prese...	9	Mus musculus	2009-03-18
+ E-TABM-667	Transcription profiling of mouse embryonic stem cell line CGR8 treated with Le...	30	Mus musculus	2009-03-18
+ E-TABM-632	Transcription profiling of human acute myeloid leukemia cells before and early ...	24	Homo sapiens	2009-02-03
+ E-TABM-628	MicroRNA profiling of human chronic lymphocytic leukemia cells in reponse to t...	12	Homo sapiens	2009-01-27
+ E-TABM-431	Chromatin immunoprecipitation of mouse hematopoietic cell lines and tissue-ty...	25	Mus musculus	2008-03-26
+ E-TABM-429	MicroRNA profiling of human acute myeloid leukemia samples from patients cha...	85	Homo sapiens	2008-03-13
+ E-TABM-405	MicroRNA profiling of patients with acute myeloid leukemia to identify microRN...	176	Homo sapiens	2008-01-09
+ E-TABM-391	Transcription profiling of human CD4+ leukemia Jurkat T-cells in which the PI3...	4	Homo sapiens	2007-12-18
+ E-TABM-346	Transcription profiling of human patients with diffuse large B-cell lymphoma tre...	53	Homo sapiens	2007-10-16
+ E-TABM-293	Comparative genomic hybridization of human chronic lymphocytic leukemia sa...	2	Homo sapiens	2007-07-17
+ E-SMDB-2850	Transcription profiling of acute myeloid leukemia FLT3 wild type and mutants in childhood AML samples from the Pediatric Oncology Group Study 9421	87	Homo sapiens	2005-12-08

Description This data set was used to study FLT3 wild type and mutants in childhood AML samples from the Pediatric Oncology Group Study 9421. and published in the Journal Blood in 2002 by Lacayo NJ et al.

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

EFO in Applications

ATLAS

[home](#) | [about the project](#) | [faq](#) | [feedback](#) | [blog](#) | [das](#) | [api](#) **net**

Genes: Organism: Conditions: View: Heatmap List

e.g. ASPM, "p53 binding" *e.g. liver, cancer, diabetes* [advanced search](#)

« 1 2 3 4 5 ... 14 » Genes 1-100 of 1364 total found (you can refine your query) • [Download all results](#) • [JSON](#) [XML](#)

Legend:  - number of studies the gene is **over/under** expressed in

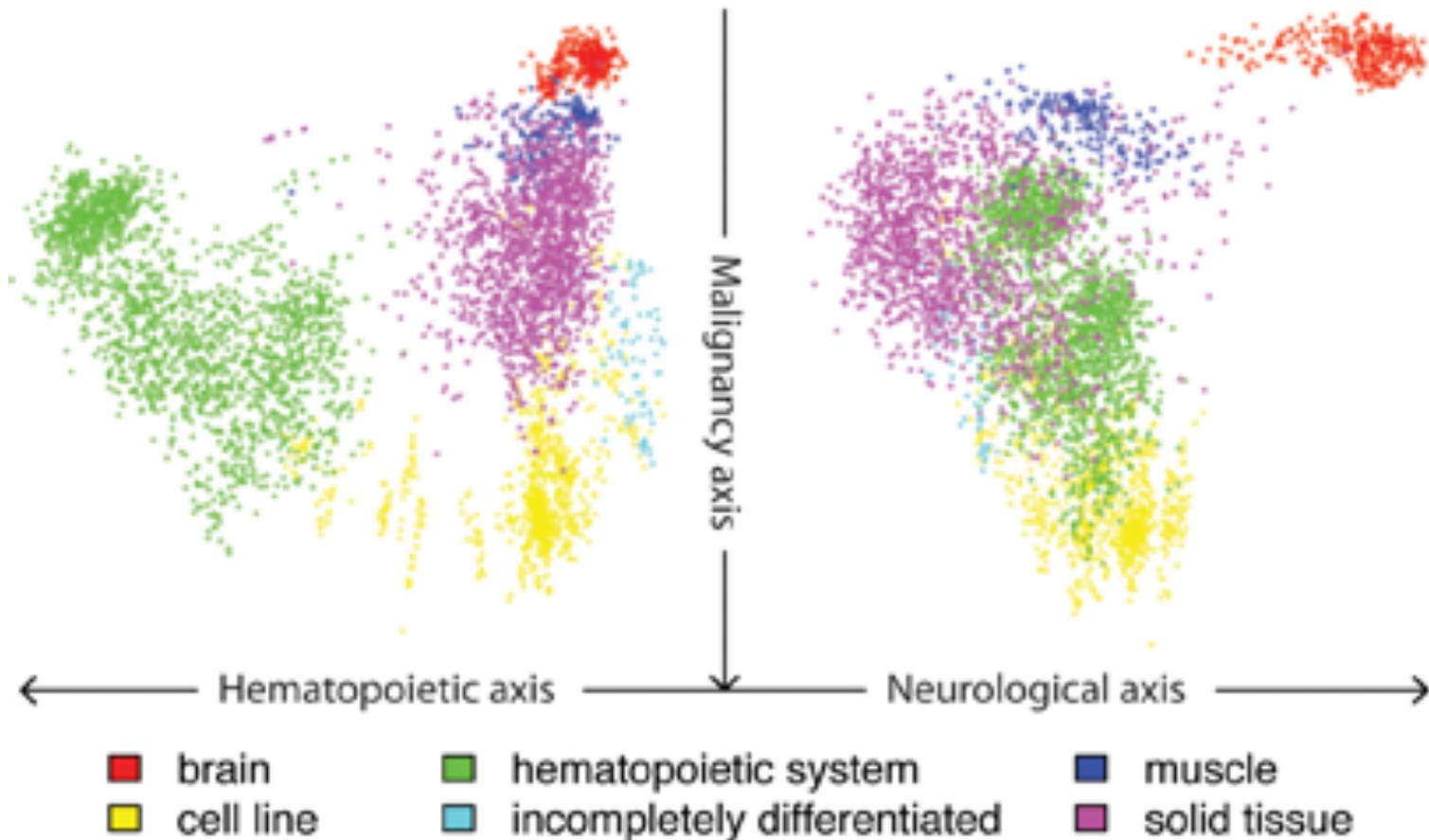
Gene	Organism	Ontology																														
		cell type	hematopoietic cell	B cell	mature B cell	pre-B cell	T cell	blood cell	fetal blood	platelet	circulating	dendritic	dendritic	erythrocyte	erythrocyte	granulocyte	neutrophil	hematopoietic stem cell	leukocyte	lymphocyte	natural killer cell	thymocyte	DP thymocyte	macrophage	memory T cell	monocyte	monocyte-derived interm...	mononuclear cell	myeloid cell	pericyte	plasma cell	plasmacytoid d
MEF2C	Homo sapiens	35/37	24/23	8/1	1	7	1	1	1	1	4/3	2	2	3	1	2	2	2	2	1	1	2	3	8	1	3	1	1	2	1	1	
MAF	Homo sapiens	38/34	26/21	7	2	8	3	1	1	1	3/2	1	4	1	2	2	4	5	2	5	1	4	1	1	4	5	3	1	1	3		
FAM65B	Homo sapiens	39/35	28/18	7	2	2	11	2	1	1	1	1	1	1	5	3	2	3	1	3	6	6	4	1	2	4	7	2	1	1	3	2

Sensible faceted browsing

Data analysis

Slims for 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

Category	Gene	Species	Cell Type	Condition	Expression	Source	Transcript Type
▶ Barrier cell (1)	TGFB1	Human	Kidney	diabetic nephropathy (severe)	Up	Rooney_FEBS Lett. 2011	mRNA
▶ Kidney cell (1)	TGFB1	Human	Kidney	Acute renal allograft rejection	Down	Head_GSE1563	mRNA
▶ Lower urinary tract cell (1)	TGFB1	Human (adult)	Glomerulus	Healthy	Present	HKUPP_Glomerulus_database	Protein
▼ Mesenchymal cell (1)	TGFB1	Human (adult)	Glomerulus	Healthy	Present	Miyamoto_J Proteome Res. 2007	Protein
▶ Smooth muscle cell (1)	Tgfb1	Mouse (adult)	Ureter	Healthy	Absent	Gudmap_GSE7887	mRNA
▶ Mesangial cell (1)	Tgfb1	Mouse (neonatal)	Ureter	Healthy	Absent	Gudmap_GSE6963	mRNA
▶ Contractile cell (1)	Tgfb1	Mouse (neonatal)	Urinary bladder	Healthy	Absent	Gudmap_GSE6466	mRNA
▶ Electrically active cell (1)	Tgfb1	Mouse (adult)	Urinary bladder	Healthy	Absent	Gudmap_GSE6466	mRNA
▶ Secretory cell (1)	Tgfb1	Mouse (neonatal)	Urinary bladder	Healthy	Weak	Gudmap_GSE4816	mRNA
▶ Motile cell (1)	Tgfb1	Mouse (adult)	Glomerular capillary endothelial cell	Healthy	Medium	Gudmap_GSE22561	mRNA
▶ Defensive cell (1)	Tgfb1	Mouse (adult)	Endothelial cell	Healthy	Medium	Gudmap_GSE22464	mRNA
▶ Leucocyte (1)	Tgfb1	Mouse (adult)	Kidney cortex cell	Healthy	Medium	Gudmap_GSE22464	mRNA
▼ Anatomy (3)							
▶ Pelvis (1)	Tgfb1	Mouse (adult)	Glomerular mesangial cell	Healthy	Medium	Gudmap_GSE20687	mRNA
▶ Lower urinary tract (1)							
▶ Kidney (3)							
▶ Loop of Henle, outer medullary portion (1)							
▼ Disease/Model (2)							
▶ Animal model (1)							
▶ Urinary system disease (1)							
▼ Biological process (3)							
▶ Biological regulation (3)							
▶ Metabolic process (3)							

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

The image shows two side-by-side screenshots from a software interface. The left screenshot, titled 'Class hierarchy (inferred): 'experimental factor'', displays a tree view of classes. The root is 'Thing', and the selected class is 'experimental factor'. Under 'experimental factor', there are several sub-classes, including 'factor_type', 'information entity', 'clinical history', 'data file', 'data format specification', 'data item', 'data set', 'discretized differential expression', 'environmental history', 'epigenetic factor', 'experiment accession', 'gene name', 'image', 'instrument design', 'label', and 'measurement'. The right screenshot, titled 'Class Annotations: 'experimental factor'', shows a list of annotations for the selected class. The annotations are: 'definition_editor' with value 'James Malone'^^string, 'definition_editor' with value 'Jie Zheng'^^string, 'definition_editor' with value 'Tomasz Adamusiak'^^string, 'description' with value 'An experimental factor in Array Express.'^^string, 'gwas_trait' with value 'true', 'label' with value 'experimental factor'^^string, and 'organizational_class' with value 'true'^^string. The 'gwas_trait' and 'organizational_class' annotations are circled in red.

EFO view

- Hierarchical relationships

The screenshot displays the EFO interface with two main panels. The left panel, titled 'Class hierarchy (inferred): forebrain', shows a tree structure of classes. The right panel, titled 'Annotations: forebrain', provides a detailed view of the 'forebrain' class, including its description, equivalent classes, superclasses, and inherited anonymous classes.

Class hierarchy (inferred): forebrain

- 'nervous system'
 - 'central nervous system'
 - brain
 - ▶ 'brain stem'
 - ▶ 'brain ventricle'
 - 'cerebral peduncle'
 - ▶ 'cingulate cortex'
 - 'corpus callosum'
 - 'corpus striatum'
 - 'corpus uteri'
 - ▶ 'cranial ganglion'
 - ▶ **forebrain**
 - amygdala
 - ▶ 'basal ganglion'
 - ▶ 'cerebral cortex'
 - diencephalon
 - epithalamus
 - ▶ 'hippocampus'
 - ▶ 'hypothalamus'
 - 'olfactory lobe'
 - ▶ 'telencephalon'
 - ▶ 'thalamus'
 - meninges

Annotations: forebrain

Annotations +

Description: forebrain

Equivalent classes +

- **part_of some forebrain**

Superclasses +

- brain

Inherited anonymous classes

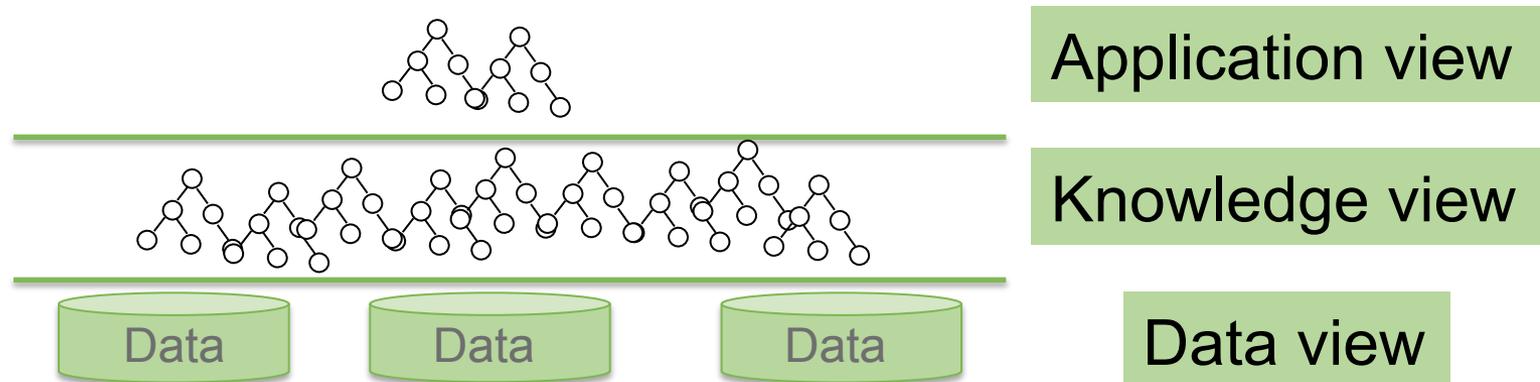
- ~~'spinal cord' or brain~~
- **has_flag some 'factor type flag'**

Members +

Keys +

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

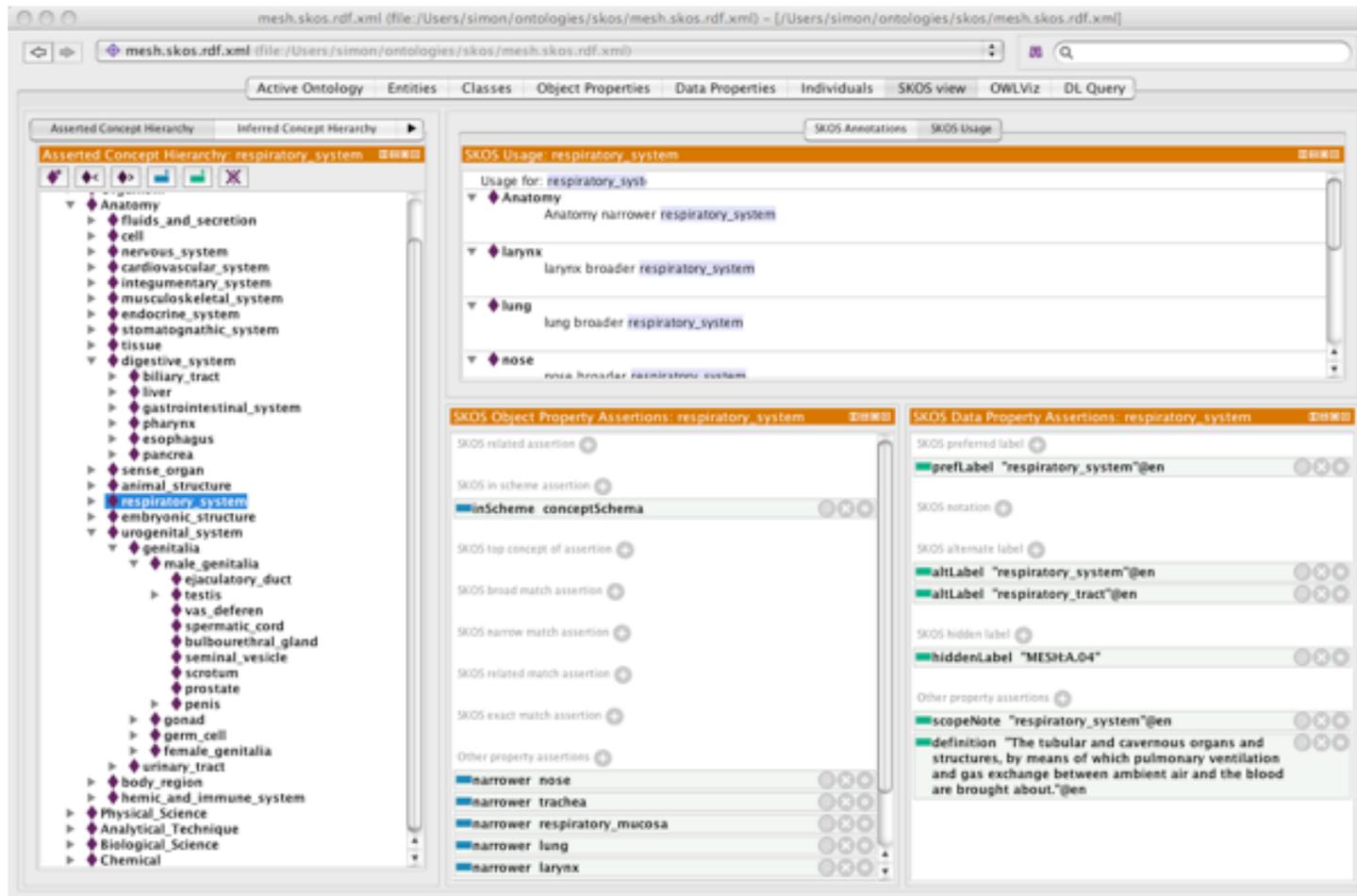


Exploring SKOS

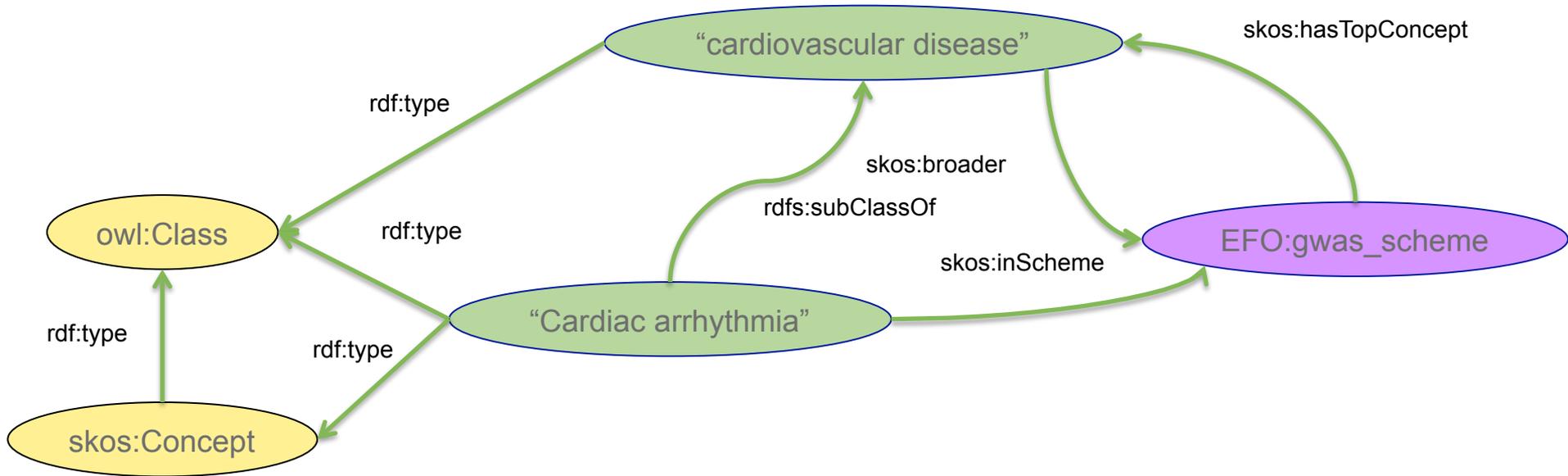
- Simple Knowledge Organization System
 - Simple W3C vocabulary for thesauri, classification schemes, taxonomies, subject heading
- Using new OWL 2 features – punning
 - A way to index terms in our ontology
 - We can safely mix OWL 2 and SKOS
 - Meta-modeling with classes and individuals
- EFO Requirements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



SKOSify EFO



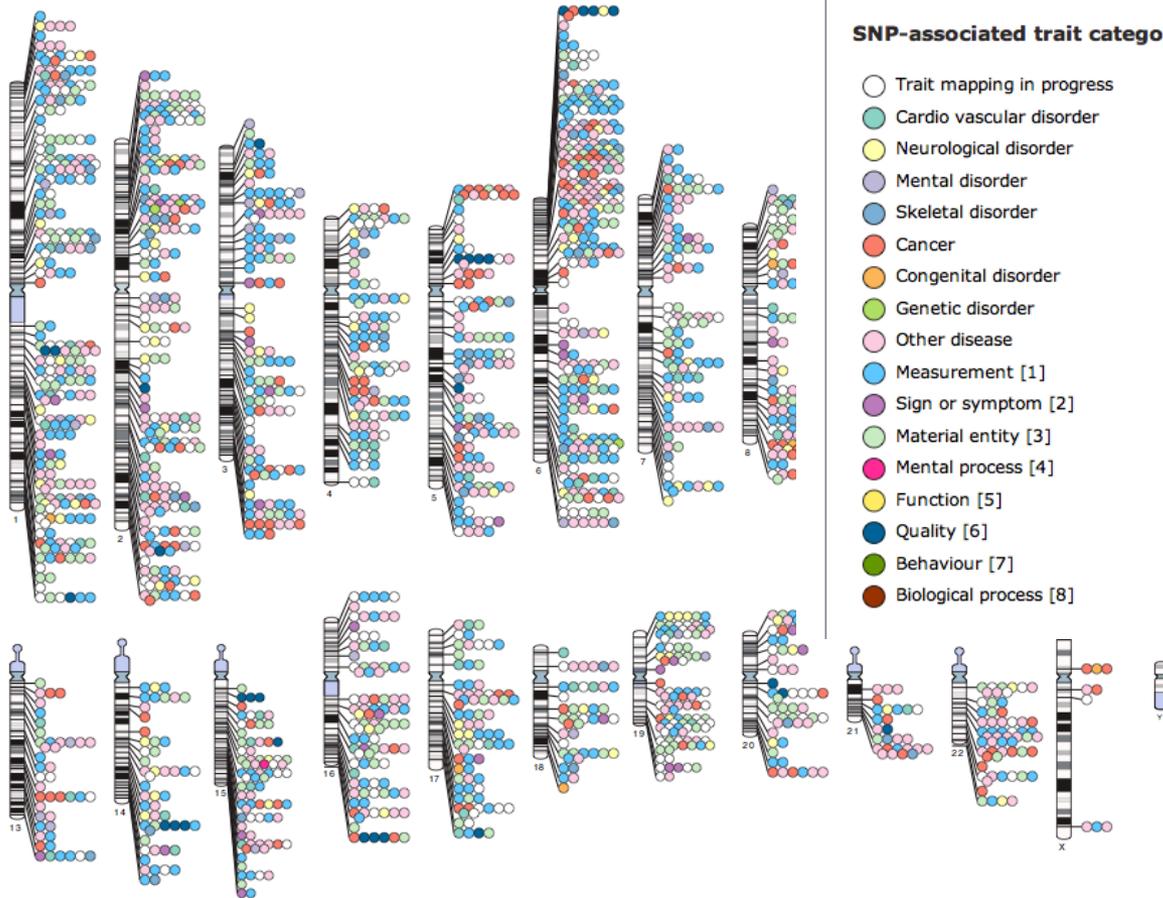
NHGRI GWAS diagram generation

GWAS Diagram Browser

Exploring Genome-wide Association Studies

GWAS Diagram Key Help About Contact

This diagram shows all SNP-trait associations with a p-value smaller than 5×10^{-8} , published in the catalogue up



GWAS Diagram Browser

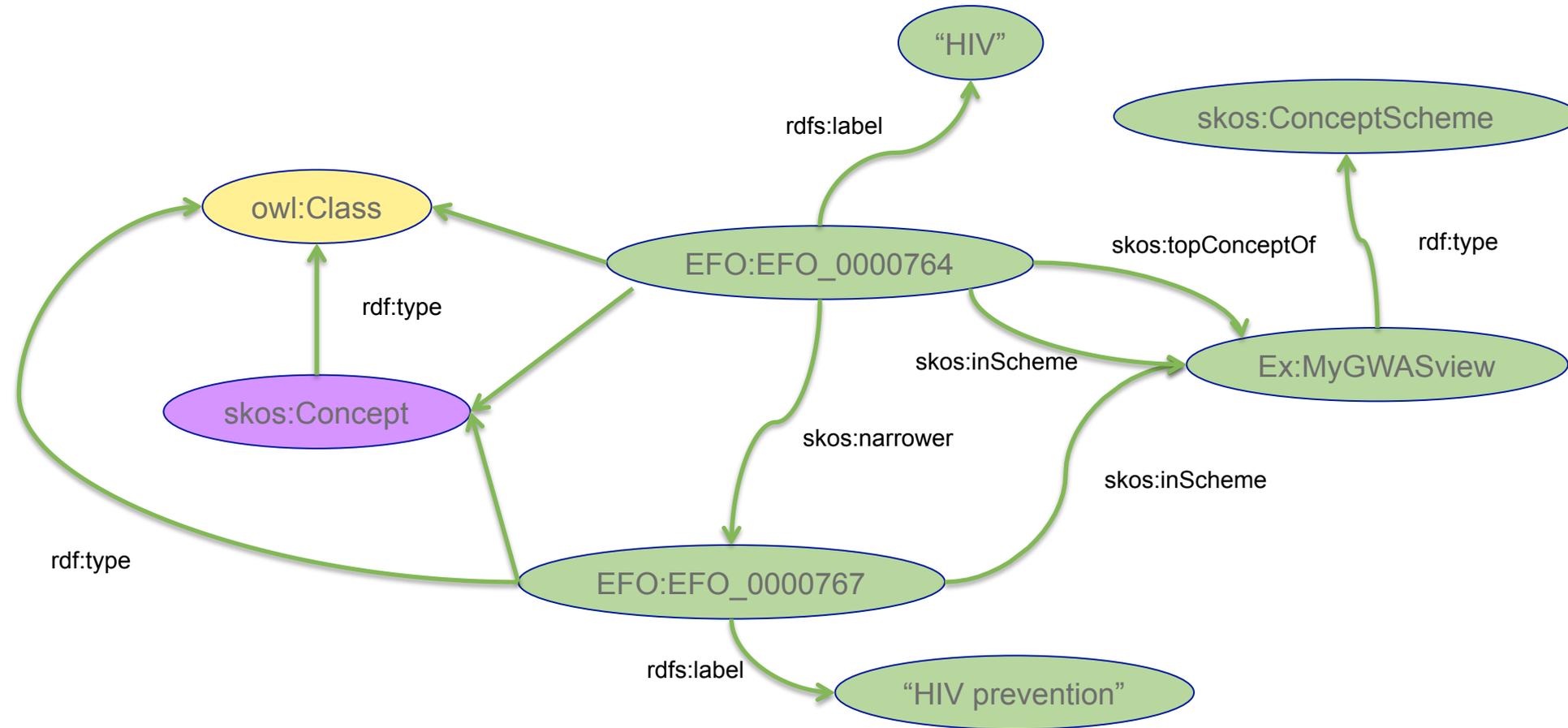
Exploring Genome-wide Association Studies

GWAS Diagram Key Help About Contact

SNP-associated trait categories

- Trait mapping in progress
- Cardio vascular disorder
- Neurological disorder
- Mental disorder
- Skeletal disorder
- Cancer
- Congenital disorder
- Genetic disorder
- Other disease
- Measurement [1]
- Sign or symptom [2]
- Material entity [3]
- Mental process [4]
- Function [5]
- Quality [6]
- Behaviour [7]
- Biological process [8]

Concept Schemes



Class hierarchy: 'experimental factor'

- Thing
 - Concept
 - ConceptScheme
 - 'experimental factor'
 - factor_type
 - 'information entity'
 - 'material entity'
 - 'material property'
 - process
 - site
 - 'obsolete class'

GWAS Diagram Browser

Exploring Genome-wide Association Studies

GWAS Diagram | Key | Help | About | Contact

SNP-associated trait categories

- Trait mapping in progress
- Cardio vascular disorder
- Neurological disorder
- Mental disorder
- Skeletal disorder
- Cancer
- Congenital disorder
- Genetic disorder
- Other disease
- Measurement [1]
- Sign or symptom [2]
- Material entity [3]
- Mental process [4]
- Function [5]
- Quality [6]
- Behaviour [7]
- Biological process [8]

Asserted Concept Hierarchy | Top Concepts Hierarchy View

Top Concepts Hierarchy View: gwas_scheme

- gwas_scheme
 - 'animal component'
 - behavior
 - 'biological process'
 - cancer
 - 'cardiovascular disease'
 - 'chemical compound'
 - 'chemical role'
 - 'congenital abnormality'
 - 'experimental factor'
 - 'genetic disorder'
 - measurement
 - 'mental or behavioural disorder'
 - 'nervous system disease'
 - 'sign or symptom'
 - 'skeletal system disease'

Punning

http://www.ebi.ac.uk/efo/EFO_0004269

Active Ontology Entities Classes Object Properties Data Properties Individuals SKOS view DL Query OPPL Patterns DualView

Class hierarchy: 'cardiac arrhythmia'

- material property'
- disposition
 - disease
 - 'acute lung injury'
 - albuminuria
 - anemia
 - 'aphthous ulcer'
 - arthrogryposis
 - 'atrial flutter'
 - 'bronchopulmonary dysplasia'
 - 'cardiovascular disease'
 - 'Diamond-Blackfan anemia'
 - 'acute hypotension'
 - 'aortic aneurysm'
 - 'aortic stenosis'
 - atherosclerosis
 - 'bundle branch block'
 - 'cardiac arrhythmia'
 - cardiomyopathy
 - 'carotid artery disease'
 - 'cerebrovascular disorder'
 - 'coronary heart disease'
 - endocarditis
 - epistaxis
 - 'heart disease'
 - 'heart failure'
 - 'heart transplant rejection'
 - hypertension
 - 'intracranial hemorrhage NOS'
 - ischemia
 - 'ischemia reperfusion injury'
 - 'myocardial infarction'
 - 'portal hypertension'
 - preeclampsia
 - 'pulmonary embolism'
 - 'pulmonary hypertension'
 - stroke
 - 'subarachnoid hemorrhage'
 - thrombophlebitis
 - 'vascular disease'

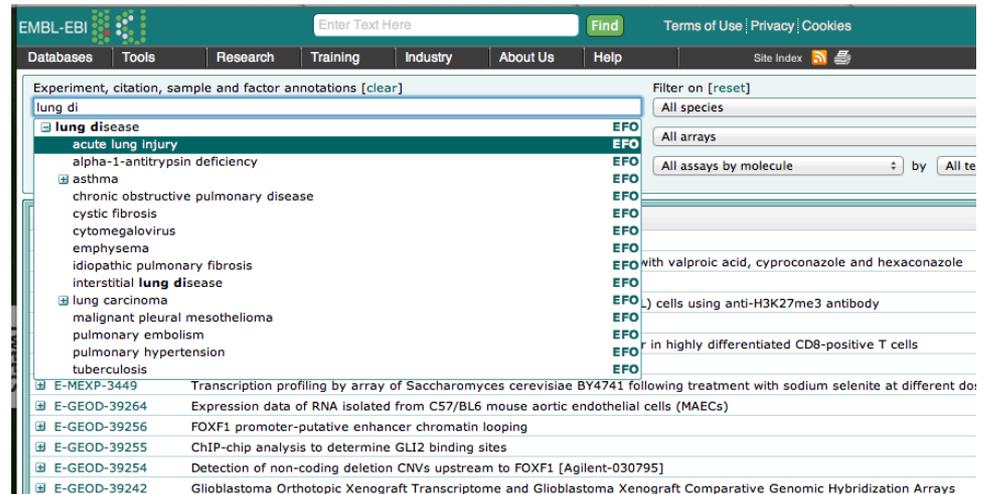
Asserted Concept Hierarchy Top Concepts Hierarchy View

Top Concepts Hierarchy View: 'cardiac arrhythmia'

- gwas_scheme
 - gwas_scheme
 - 'animal component'
 - behavior
 - 'biological process'
 - cancer
 - 'cardiovascular disease'
 - 'Abdominal Aortic Aneurysm'
 - 'Behcet's syndrome'
 - 'Moyamoya disease'
 - atherosclerosis
 - 'atrial fibrillation'
 - 'brain aneurysm'
 - 'brain infarction'
 - 'cardiac arrhythmia'
 - 'cardiac hypertrophy'
 - 'carotid artery disease'
 - 'coronary heart disease'
 - 'dilated cardiomyopathy'
 - 'heart failure'
 - hypertension
 - 'myocardial infarction'
 - stroke
 - 'subarachnoid hemorrhage'
 - 'sudden cardiac arrest'
 - 'thoracic aortic aneurysm'
 - 'vascular disease'
 - 'chemical compound'
 - 'chemical role'
 - 'congenital abnormality'
 - 'experimental factor'
 - 'genetic disorder'
 - measurement
 - 'mental or behavioural disorder'
 - 'nervous system disease'
 - 'sign or symptom'
 - 'skeletal system disease'

Viewgen framework (Under development)

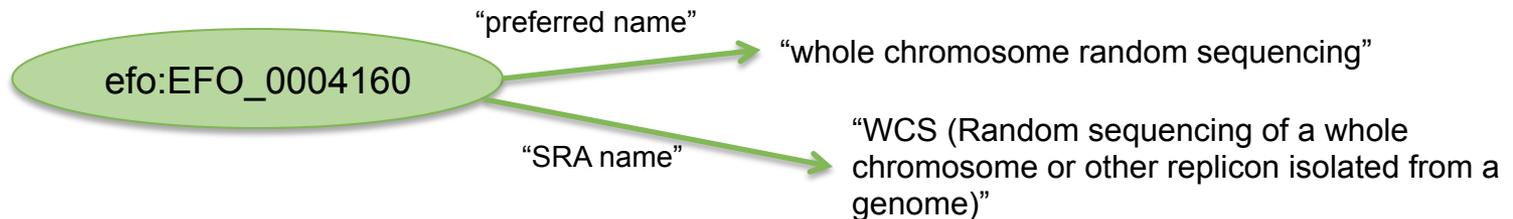
- Java API for extracting ontology views
 - Generate OWL or SKOS exports
 - OWL DL, SPARQL or Annotation based
 - Protégé 4.1 plugin
- Future work
 - Web UI (term shopping cart)
 - JQuery widget library for embedding ontology views in web application
 - Other requirements?



The screenshot shows the EMBL-EBI website interface. At the top, there is a search bar with the text "Enter Text Here" and a "Find" button. Below the search bar, there is a navigation menu with tabs for "Databases", "Tools", "Research", "Training", "Industry", "About Us", and "Help". The main content area displays a search result for "lung di". The results are organized into a list of terms, each with a checkbox and a corresponding EFO (Eukaryotic Orthology) identifier. The terms include "lung disease", "acute lung injury", "alpha-1-antitrypsin deficiency", "asthma", "chronic obstructive pulmonary disease", "cystic fibrosis", "cytomegalovirus", "emphysema", "idiopathic pulmonary fibrosis", "interstitial lung disease", "lung carcinoma", "malignant pleural mesothelioma", "pulmonary embolism", "pulmonary hypertension", and "tuberculosis". To the right of the terms, there is a "Filter on [reset]" section with options for "All species", "All arrays", and "All assays by molecule". Below the terms, there is a table of experiments, including "E-MEXP-3449" (Transcription profiling by array of Saccharomyces cerevisiae BY4741 following treatment with sodium selenite at different do...), "E-GEOD-39264" (Expression data of RNA isolated from C57/BL6 mouse aortic endothelial cells (MAECs)), "E-GEOD-39256" (FOXF1 promoter-putative enhancer chromatin looping), "E-GEOD-39255" (ChIP-chip analysis to determine GLI2 binding sites), "E-GEOD-39254" (Detection of non-coding deletion CNVs upstream to FOXF1 [Agilent-030795]), and "E-GEOD-39242" (Glioblastoma Orthotopic Xenograft Transcriptome and Glioblastoma Xenograft Comparative Genomic Hybridization Arrays).

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



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)