Examples for *OntologyTerm* use

Current schema implementation of this document

Info: Ontogenesis blog
Info: Working implementation of the GA4GH docsystem's Ontologies document

Why should we use an ontology term?

Info: http://ontogenesis.knowledgeblog.org/1296

A user may want to retrieve the rsIDs of all genomic variants to ciliopathies. Each rsID is annotated with a specific disease (e.g. Bardet Biedl syndrome, orofaciodigital syndrome). To query by a the functional grouping ‘ciliopathy’, classification of these diseases ciliopathies is needed, and can be provided through an ontology.

Non-ontology example:

- **rs587777067** is associated with orofaciodigital syndrome (www.ensembl.org/….s587777067)
- a query for ciliopathy (or ciliopathies) returns two different results in the current Ensembl representation
  - http://www.ensembl.org/Multi/Search/Results?q=ciliopathies;site=ensembl_all
  - http://www.ensembl.org/Multi/Search/Results?q=ciliopathy
- neither of these represents all known ciliopathies or returns rs587777067, as these queries are not performing a lookup for child terms of a functional class ciliopathy, but rather represent text queries for the two lexical variants of the term

Ontology Lookup Service

- http://www.ebi.ac.uk/efo/EFO_0003900
- multiple child classes are returned, including those without a lexical match to the disease name

The effective use of ontology lookups requires the annotation of rsIDs with unique identifiers for the associated diseases, so that a programmatic lookup can use these to identify their parents and/or relations. Text queries are likely to return partial or erroneous result sets. Ontologies overlap in their scope, design and content. In the case of results from different ontologies, which may have a varying depth, the executioner of the query has to judge about the optimal scope of the returned data.

What is the minimum attribute requirement for OntologyTerm in GA4GH?

Conceptually (and consistent with the metadata branch)

ontologyId	required and implemented as URI we assume this resolves to a meaningful document, e.g. http://purl.obolibrary.org/obo/SO_0000147
term	preferred but not required (e.g. ‘exon’); corresponds to class label

sourceName not required since should be resolved from prefix etc., but supporting/fall-back in case of non-standard/deprecated/entropic annotations; possible use for CURIEs in compact sequence ontology implementations (e.g. SO:0000147)
sourceVersion not required but good practice; if no explicit versioning, ISO8601 formatted data of retrieval should be used

Ontology Selection and Overlap

Sometimes a single ontology provides excellent coverage of a domain. For example the Sequence Ontology is used successfully in GFF files to annotate exons, introns etc. There are multiple ontologies in some domains which overlap in scope and content and also interoperation between ontologies. For example the Human Phenotype Ontology (HP) provides terms describing human phenotypes. Disease phenotype associations are not provided in the HP, rather as supporting files with common cross references such as OMIM identifiers. When selecting an ontology consider coverage - how much of your data is represented,=: structure - does the ontology provide structure that meets your use cases, e.g. contains a class ciliopathy (see above), update frequency, ability to request terms when needed, adherence to community standards e.g. OBO foundry provides recommendations on versioning strategy and term deprecation. Note that OBO policy dictates that when the meaning of a class changes, then the identifier/IRI is deprecated/obsoleted, and a new identifier/IRI is minted. As a consequence, many databases that store associations to OBO classes (genes, diseases to phenotype etc) do not record the version of the ontology, as the semantics of the ID can be treated as immutable.

Info: Why use the Human Phenotype Ontology (blog post by MH)

Age, date, time interval values => ISO8601

Examples

Genotypic sex:

ontologyId: "http://purl.obolibrary.org/obo/PATO_0020001”,
term: "male genotypic sex" ,
sourceName: "PATO Phenotypic quality",

Sequence Ontology:

ontologyId: "http://purl.obolibrary.org/obo/SO_0001583”,
term: "missense_variant",
sourceName: "Sequence Ontology",
sourceVersion:"release_2.5.3"

Human Phenotype ontology

ontologyId: “http://purl.obolibrary.org/obo/Diabetes mellitus”,
term: “Diabetes mellitus,
sourceName: "human_phenotype Ontology",
sourceVersion: "release_Jan2016*"

ontologyId: “http://www.ebi.ac.uk/efo/HP_0012059”,
term: “Lentigo maligna melanoma”,

sourceName: “human_phenotype_ontology",
sourceVersion: “2016-01-14”

Body part (Uberon)

ontologyId: “http://www.ebi.ac.uk/efo/UBERON_0003403”,
term: “skin of forearm”,
sourceName: “uberon”,
sourceVersion: “2015-11-23”

Human disease ontology

ontologyId: “http://purl.obolibrary.org/obo/DOID_9351”,
term: “diabetes mellitus”,
sourceName: "disease_ontology",
sourceVersion: "2016-01-25"

Experimental factor ontology

ontologyId: “http://purl.obolibrary.org/obo/EFO_0000400”,
term: “diabetes mellitus”,
sourceName: “experimental_factor_ontology",
sourceVersion: “V2.68”

ontologyId: “http://www.ebi.ac.uk/efo/EFO_0004422”,
term: “exome”,
sourceName: "Experimental Factor Ontology",
sourceVersion: "release_2.68"

SNOMEDCT representation of ICD-O 3 Cancer Histology

ontologyId: “http://purl.bioontology.org/ontology/SNMI/M-94703“
term: “Medulloblastoma, NOS”
sourceName: “SNOMED CT model component”
sourceVersion:“2016-01-28”

Unit Ontology

ontologyId: “http://purl.obolibrary.org/obo/UO_0000016”,
value: “4”,
term: “millimetre”,
sourceName: "Unit Ontology",
sourceVersion: "2015-12-17"

The discussion of the concrete use of OntologyTerm objects to annotate “biofeatures” has been moved to the BioData object development page

Lint

Example:

A representation is to mint a measurement class - in this case a similar scenario has been modelled for GWAS data for simplicity of query, curation and visualisation

ontologyId: “http://www.ebi.ac.uk/efo/EFO_0007009”
term: “skin pigmentation measurement”
sourceName: “Experimental Factor Ontology”
sourceVersion: “2.68”

This from Tony Brookes:

I propose we also need the options of

- including one or modifiers (e.g., "age of onset", "average value", "left eye", etc)
- having multiple values per phenotype, each with its own unit
- expressing modifiers or values or units as ontology entries

Therefore:

```
 phenotypeTerm - required, string
 sourceName - optional, string
 sourceVersion - optional, string
 id - optional, string
 URI - optional, string, string or ontology entry
 modifiers - optional, string
 value - optional, string or ontology entry
 unit - optional, string or ontology entry
```

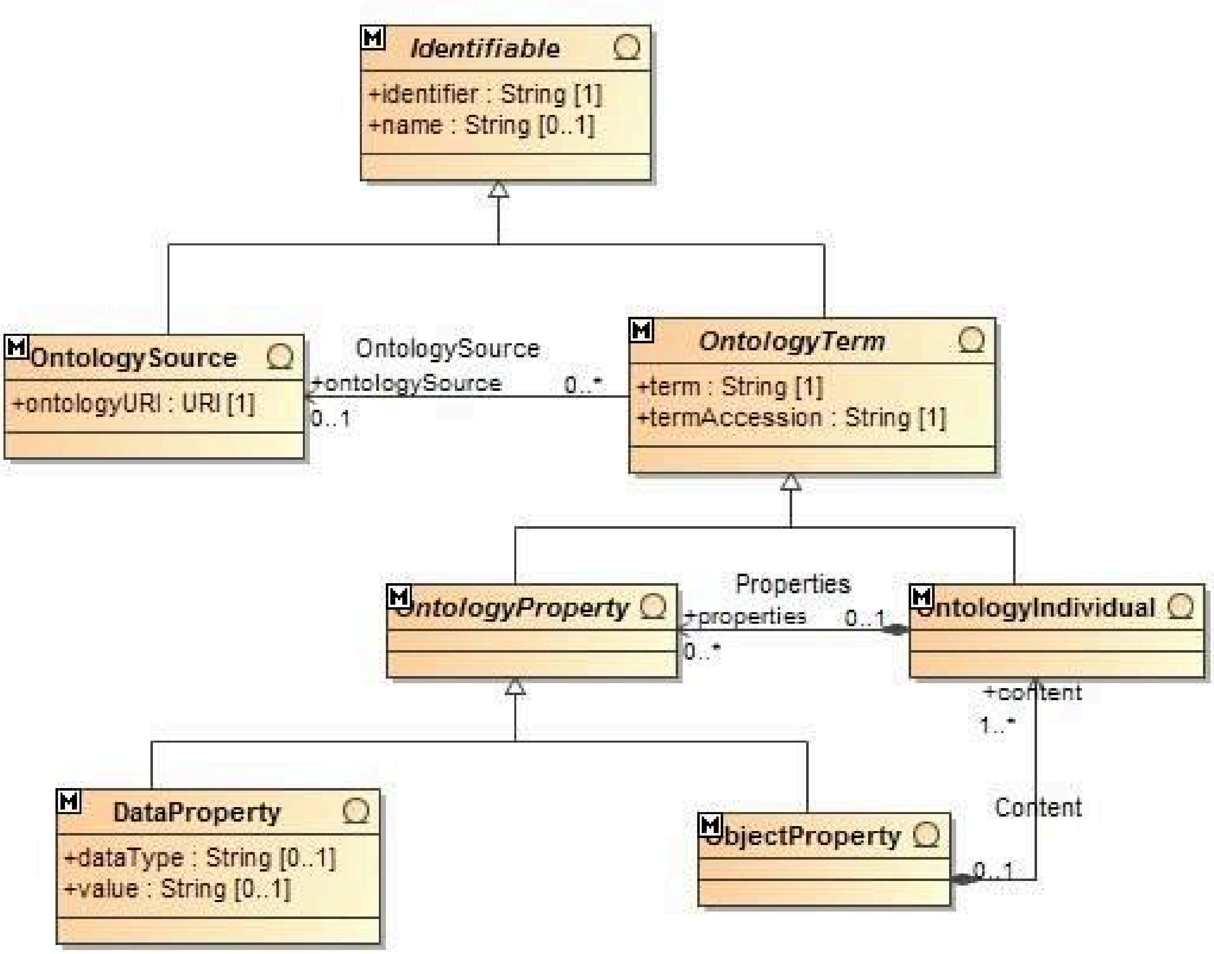
Organised as:

```
phenotypeTerm {sourceName, sourceVersion, id, URI, [{modifier1},
{modifier2},...], [{{value1}:{unit1}}, {{value2}:{unit2}},...]}
```

E.g.,

```
* 4mm skinpunch biopsy to describe a sample source
“skinpunch biopsy” {"eagle-i Research Resource Ontology",
"2013-08-02”, “ERO_0001336”,
“http://www.ebi.ac.uk/efo/ERO_0001336”, , {"4":“millimetre”}}
```

* Number of occurrences of lentigo melanoma on specified area of aged skin
“Lentigo maligna melanoma” {“human_phenotype_ontology", “14-01-2016”, “HP_0012059”, “http://www.ebi.ac.uk/efo/HP_0012059”, “skin of forearm” {“uberon”, “23-11-2015”, “UBERON_0003403”, “http://www.ebi.ac.uk/efo/UBERON_0003403”, , }, {“10”:“square centimetre” {"Unit Ontology", "17-12-2015", “UO_0000081”, “http://purl.obolibrary.org/obo/UO_0000081”, , }}}



note that the simple case is covered but allows nested property definitions which would cover ‘units’

Question(Tony) - most of this is self-apparent, but what is the ontologyIndividual concept designed to represent?
Answer(mm) - Individuals in ontologies is a debated question as to whether there should be any. for terminologies, they are all probably OntologyTerms not an Individual, but there is a debate whether describing phenotypes or alleles, for instance, for a patient, whether those should be treated as the specific instances associated with that patient or assigning simply assigning a general term. for ga4gh, this distinction is likely not important and OntologyIndividual can go away and the ObjectProperty content association can go directly to OntologyTerm.

Use case

Eliminate batch effects in multiple runs of an Affy array investigating expression profiles in breast cancer cells

Job story
When running a high throughput technology (NGS, microarray etc) to answer a biological question, as the scientist needing to understand the biological conclusions I want know the date, source of reagent, hardware etc used for the study/assay in order to eliminate the batch effects of running multiple batches of samples and identify the non-biological variation that can increase error.

Ontology term
“affymetrix array” {"eagle-i Research Resource Ontology", "2013-08-02”, “ERO_0001265”, “http://www.ebi.ac.uk/efo/ERO_0001265”, , {
Reagents: GeneChip® WT PLUS Sufficient for 10 reactions Reagent Kit
Instrument: GeneChip® Human Transcriptome Array 2.0
Number of samples: 30
Date of run: 2015-04-01
}