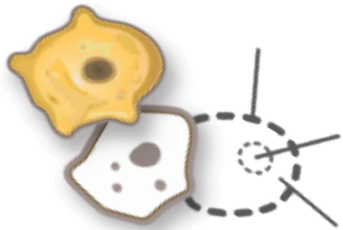
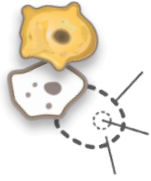


# Ontology-based Tools to Enhance the Curation Workflow

**Trish Whetzel**  
**Outreach Coordinator**



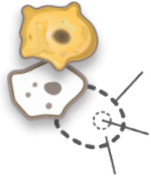
THE NATIONAL CENTER FOR  
**BIOMEDICAL ONTOLOGY**



# Enhancing the Curation Workflow

- Curation
  - The activity of organizing, representing and making biological information accessible to both humans and computers<sup>1</sup>
- Constraints
  - Large amounts of data
  - Time-consuming
- Components to enhance
  - Data submission
  - Ontology enrichment
  - Annotation of textual metadata

<sup>1</sup>Howe *et al.* [Nature. 2008 Oct 2;455\(7213\):590](#)



# BioPortal

**BioPortal** Browse Search Projects Annotate All Mappings All Resources Alpha Logged In As whetzel [Help/About](#) [Send Feedback](#)  
[My Account](#) [Log Out](#)

Biomedical Resource Ontology Gene Ontology Extension Experimental Factor Ontology Pathway ontology Ontology for Biomedical Investigations

## Welcome to NCBO BioPortal

Use BioPortal to access and share ontologies that are actively used in biomedical communities. You can search for terms in ontologies (try typing "Melanoma" in the "Search all ontologies" box in the left column), browse a list of ontologies in BioPortal (type "NCI Thesaurus" in the "Find an ontology" box in the middle column), search biomedical resources that we automatically annotated with ontology terms (try typing "Melanoma" in the "Search resources" box in the right column). You can [create ontology-based annotations for your own text](#), [link your own project that uses ontologies to the description of those ontologies](#), [find and create relations between terms in different ontologies](#), [review and comment on ontologies and their components as you browse them](#). [Sign in to BioPortal](#) to submit a new ontology or ontology-based project, provide comments on ontologies or add ontology mappings.

Search all ontologies

  
[Advanced Search](#)

Find an ontology

  
[Browse Ontologies >](#)

Search resources

  
[Advanced Resource Search](#)

Most Viewed Ontologies (August, 2010)

Ontology	Views
<a href="#">SNOMED Clinical Terms</a>	2044
<a href="#">NCI Thesaurus</a>	1294
<a href="#">NCBI organismal classification</a>	1177
<a href="#">RadLex</a>	738
<a href="#">MedDRA</a>	639

Statistics

Ontologies	219
Terms	1,438,792
Resources Indexed	22

Latest Notes

[Order Chiroptera has been included \(Malaria Ontology\)](#) 20 days ago by topalis  
Bats are now correctly appearing as members of Chiroptera.

[New Relationship Proposal: is\\_a \(Malaria Ontology\)](#) 29 days ago by slozano

[New Term Proposal: ventral fin lepidotrichium \(Teleost Anatomy Ontology\)](#) about 1 month ago by hlapp

[Spelling error \(Chemical entities of biological interest\)](#) 2 months ago by memartone  
Acetylcholine is spelled wrong in the definition

[relationship with Smoking Behavior \(NCI Thesaurus\)](#) 4 months ago by pkcourtney  
Seems odd that Tobacco Use as a term is

Latest Mappings

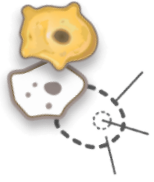
[Patient \(Health Level Seven\) => Intervention \(SNOMED Clinical Terms\)](#)  
08/06/10 nigam

[Melanoma \(NCI Thesaurus\) => Malignant melanoma \(SNOMED Clinical Terms\)](#)  
08/02/10 nigam

[Malignant melanoma \(SNOMED Clinical Terms\) => Melanoma \(NCI Thesaurus\)](#)  
08/02/10 nigam

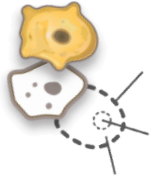
[JERMOntology:BioMaterial \(SysMO-JERM\) => BioMaterial \(MGED Ontology\)](#)  
BioPortal UI07/07/10 kwoilstencroft

<http://bioportal.bioontology.org/>



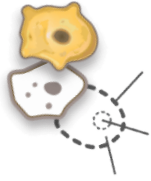
# Enhancing the Curation Workflow

- Data submission
  - Ontology Widgets and Ontology Web services
- Ontology enrichment
  - BioPortal Notes
- Annotation of textual metadata
  - NCBO Annotator



# Tools for Data Submission

- Ontology Widgets
  - Code to use in your web site
  - Search for and select terms from ontologies to annotate your data
  - Enables consistent annotation of data by re-use of the properly formatted term
  - Available for all ontologies in BioPortal



# Tools for Data Submission

BioPortal | Browse | Search | Projects | Annotate | All Mappings | All Resources Alpha

NCBI organismal classification

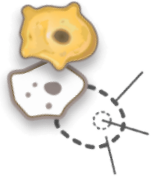
## Gene Ontology Extension [Subscribe to updates](#)

Metadata | Views (0) | Projects (0) | Reviews (0) | Notes (0) | Metrics | **Ontology Widgets**

### Adding NCBO Widgets for Gene Ontology Extension to Your Site

Step 1: Select a widget

Widget Description	Widget Demonstration
Type a term in Gene Ontology Extension and jump to it in BioPortal <a href="#">Get Code</a>	Jump To: <input type="text"/> <a href="#">Go to BioPortal</a>
Fill in a field in your form with a term from Gene Ontology Extension <a href="#">Get Code</a>	Example 1 (start typing the term name to get its full URI) <code>http://purl.bioontology.org/ontology/GO/GO_0005490</code> Example 2 (get the ID for a term) <code>p450</code> <ul style="list-style-type: none"><li>cytochrome <b>P450</b> (preferred name)</li><li>oxygen binding (synonym)</li><li>cytochrome <b>P450</b> 4A1-heme linkage (preferred name)</li><li>cytochrome <b>P450</b> activity (preferred name)</li><li>epoxygenase <b>P450</b> pathway (preferred name)</li><li>25-hydroxycholecalciferol-24-hydroxylase activity (synonym)</li><li>5beta-cholestane-3alpha,7alpha-diol 12alpha-hydroxylase activity (synonym)</li><li>arachidonic acid 11,12-epoxygenase activity (synonym)</li><li>arachidonic acid 14,15-epoxygenase activity (synonym)</li><li>arachidonic acid epoxygenase activity (synonym)</li><li>arachidonic acid monooxygenase activity (synonym)</li><li>calcidiol 1-monooxygenase activity (synonym)</li><li>cholestanetriol 26-monooxygenase activity (synonym)</li></ul> Results provided by <a href="#">NCBO BioPortal</a>
Get RSS feed for changes in Gene Ontology Extension <a href="#">Get Code</a>	<a href="#">Full Version</a>



# Ontology Widgets

- BioPortal Widgets

Jump To:

- melanoma** (preferred name)
- cutaneous **melanoma** (preferred name)
- metastatic **melanoma** (preferred name)
- na** (preferred name)

**Jump To**

Example 1 (start typing the term name to get its full URI)  
`http://www.ebi.ac.uk/efo/EFO_0000756`

Example 2 (get the ID for a term)  
`EFO_0000756`

Example 3 (get the preferred name for a term)

- melanocyte** (preferred name)
- MELAS syndrome** (preferred name)
- Drosophila **melanogaster** (preferred name)
- mela noma** (preferred name)
- Drosophila **mela**

**Auto-complete**

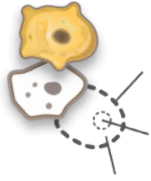
Get RSS feed for changes in Experimental Factor Ontology

[Ontology added](#) 11/05/09  
Ontology Experimental Factor Ontology version 1.7 was added to the repository

[Ontology added](#) 10/07/09  
Ontology Experimental Factor Onto

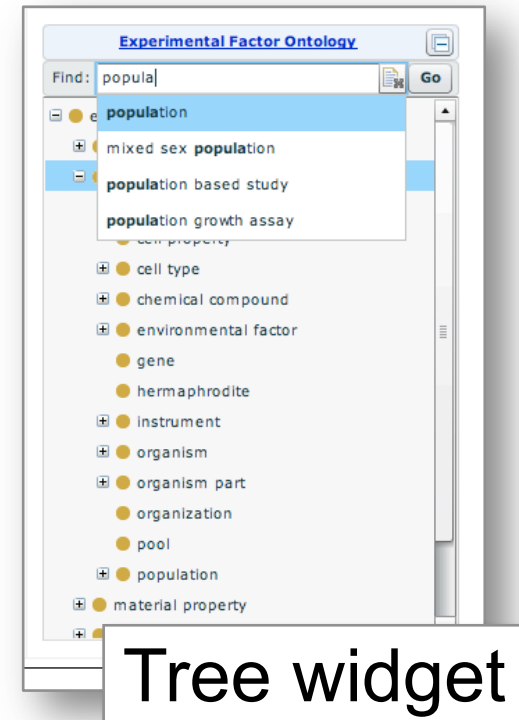
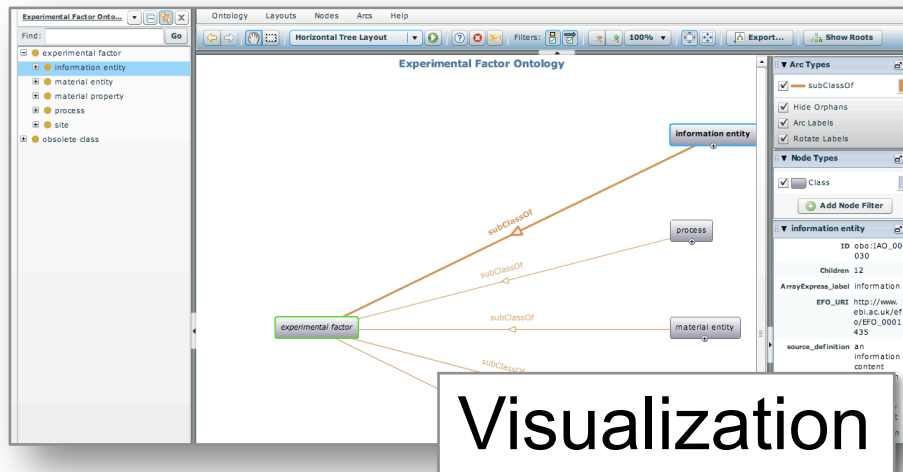
BioPortal

**RSS feed**

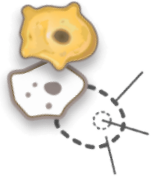


# Ontology Widgets

- BioPortal Widgets

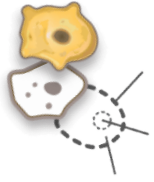






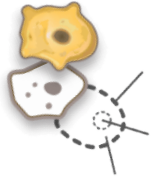
# Tools for Data Submission

- Ontology Web services
  - Access to ontology content via REST services
- Types
  - Search across all BioPortal ontologies
  - Get Term details
  - Get Term parents, children or siblings
  - Extract subsets of terms
- Available for all ontologies in BioPortal
  - [http://www.bioontology.org/wiki/index.php/BioPortal\\_REST\\_services](http://www.bioontology.org/wiki/index.php/BioPortal_REST_services)



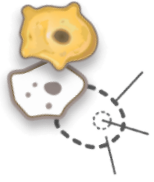
# Enhancing the Curation Workflow

- Data submission
  - Ontology Widgets and Ontology Web services
- **Ontology enrichment**
  - BioPortal Notes
- Annotation of textual metadata
  - NCBO Annotator

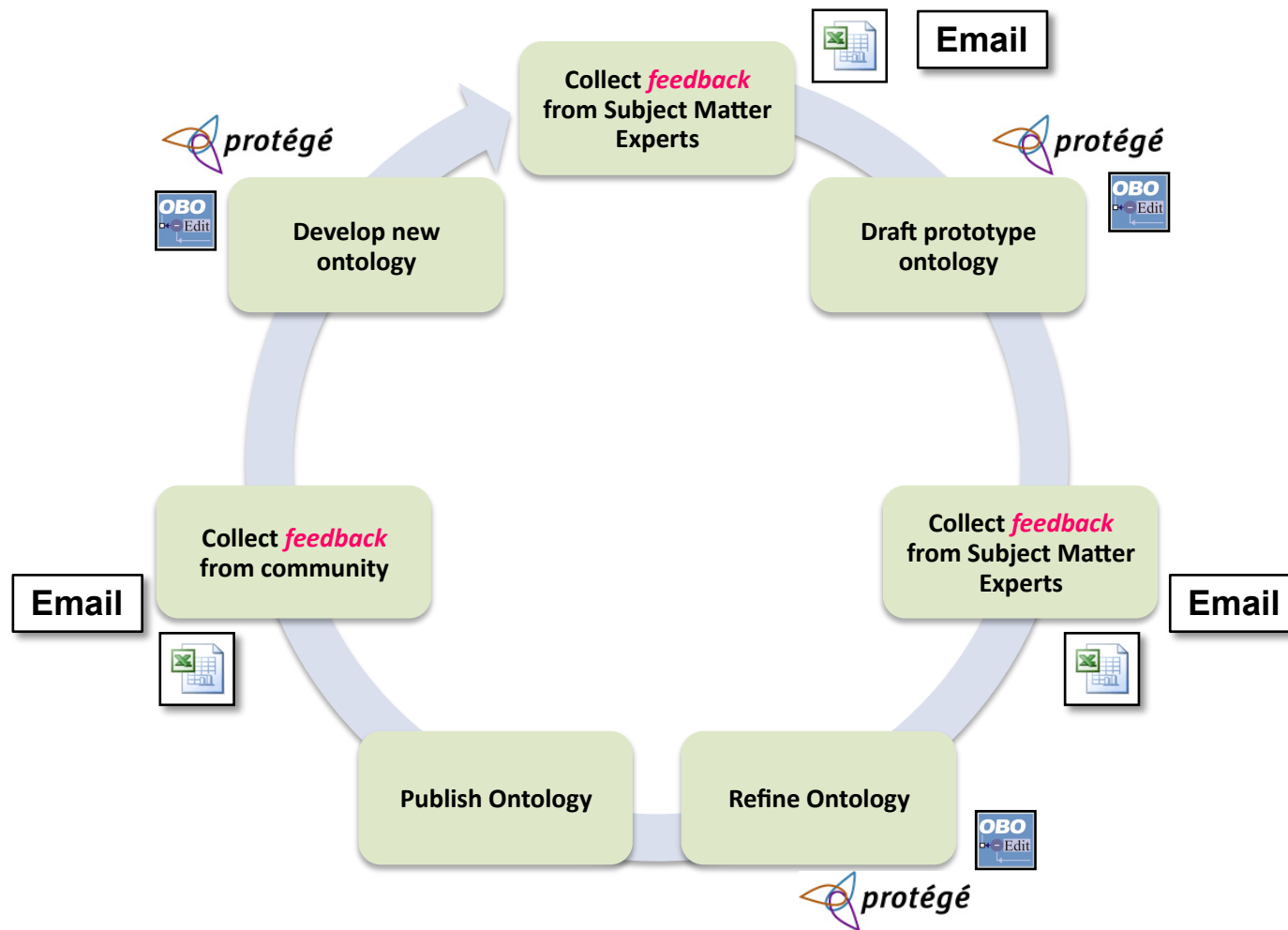


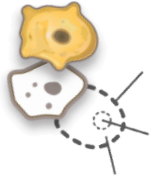
# Tools for Ontology Enrichment

- Ontology Enrichment
  - Expansion of the ontology based on user need for terms
- Constraints
  - Existing trackers do not collect structured information
  - No programmatic access to tracker
  - Lack of integration with ontology editing software

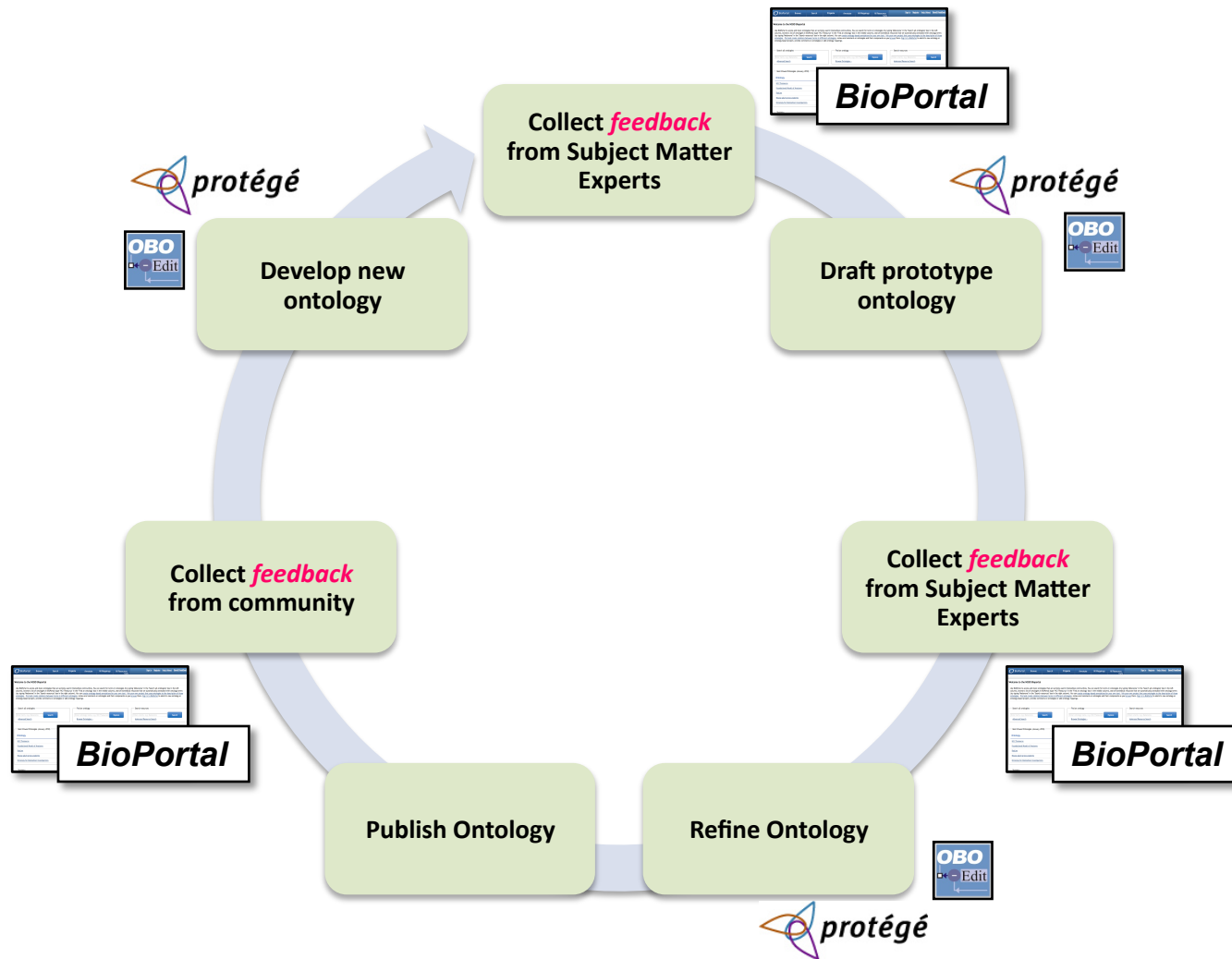


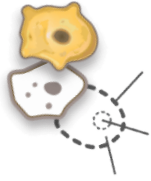
# Ontology Development Lifecycle





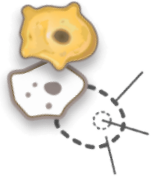
# Ontology Development Lifecycle





# BioPortal Notes

- Notes
  - Provide a mechanism to collect structured information
  - Programmatic access
  - Alerts of updates from both Email and RSS
  - Integration with ontology editing programs



# BioPortal Notes

**BioPortal** | Browse | Search | Projects | Annotate | All Mappings | All Resources Alpha | Logged In As whetzel | Help/About | Send Feedback | My Account | Log Out

Biomedical Resource Ontology | NCI Thesaurus

## Biomedical Resource Ontology

 | Subscribe to updates

[Submit New Version](#) | [Edit Ontology Information](#)

Metadata | Views (6) | Projects (4) | Reviews (1) | **Notes (66)** | Metrics | Ontology Widgets

Show 50 entries | Search:

SUBJECT	AUTHOR	TYPE	TARGET
<a href="#">New Term Proposal: Exercise Study Facility</a>	Mette	New Term Proposal	<a href="#">Physiology Facility (Class)</a>
<a href="#">New Term Proposal: Sleep Study Facility</a>	Mette	New Term Proposal	<a href="#">Physiology Facility (Class)</a>
<a href="#">Deprecate</a>	whetzel	Comment	<a href="#">Symbolic and Analytic Model (Class)</a>
<a href="#">re-activate this term</a>	whetzel	Comment	<a href="#">Web Service (Class)</a>
<a href="#">Move in hierarchy</a>	whetzel	Comment	<a href="#">Material Analysis Service (Class)</a>
<a href="#">Proposal</a>	whetzel	Comment	<a href="#">People Resource (Class)</a>
<a href="#">Add Education Service as a Synonym</a>	whetzel	Comment	<a href="#">Training Service (Class)</a>
<a href="#">Synonym proposal</a>	whetzel	Comment	<a href="#">Robotic Arrayer (Class)</a>
<a href="#">Term name</a>	whetzel	Comment	<a href="#">Molecular Modeling and Classification (Class)</a>
<a href="#">Re-structure branch</a>	whetzel	Comment	<a href="#">People Resource (Class)</a>
<a href="#">Discuss this branch with collaborating groups, e.g. eagle-i, OBI, NIF</a>	whetzel	Comment	<a href="#">Instrument (Class)</a>
<a href="#">Review definition</a>	whetzel	Comment	<a href="#">Physiology Facility (Class)</a>
<a href="#">Refine definition</a>	whetzel	Comment	<a href="#">Molecular Biology Facility (Class)</a>
<a href="#">Add Image to Data_Resource branch</a>	whetzel	Comment	<a href="#">Image (Class)</a>
<a href="#">Clarify action item</a>	whetzel	Comment	<a href="#">Programmatic Access (Class)</a>

Biomedical Resource Ontology Version 3.1 | Physiology Facility | Link Here | Subscribe

View Ontology Summary | Details | Visualization | **Notes (3)** | Mappings (0) | Resource Index

Jump To:  Go

Legend

- Activity
- Area of Research
- Biositemaps Information Model
- core:Collection
- core:Concept
- core:ConceptScheme
- Deprecated Activity
- Deprecated Area of Research
- Deprecated Resource
- Resource
  - Funding Resource
  - Information Resource
  - Material Resource
    - Assessment Material Resource
    - Biomaterial Supply Resource
    - Biomedical Supply Resource
    - Facility Core
      - Biosafety Level Facility
      - Cell Biology Facility
      - Fabrication Facility
      - Imaging Facility
      - Molecular Biology Facility**
      - Physiology Facility**
      - Research Animals Facility
      - Tissue Organ Facility
  - Instrument
    - Laboratory Supply Resource
    - Medical Device
    - Reagent Resource
  - People Resource
  - Service Resource
  - Software
  - Training Resource

Show 50 entries | Search:

SUBJECT	AUTHOR	TYPE	CREATED
<a href="#">New Term Proposal: Exercise Study Facility</a>	Mette	New Term Proposal	06/15/2010

### New Term Proposal: Exercise Study Facility

New Term Proposal submitted by Mette 3 months ago on [Physiology Facility](#) in [Biomedical Resource Ontology](#)

PREFERRED NAME	PROPOSED ID	PARENT
Exercise Study Facility		<a href="http://bioontology.org/ontologies/BiomedicalResourceOntology.owl#Physiology_Facility">http://bioontology.org/ontologies/BiomedicalResourceOntology.owl#Physiology_Facility</a>

REASON FOR CHANGE	STATUS	CONTACT INFO
Physiology facility child		

SYNONYMS

DEFINITION A facility or core devoted to exercise studies

### Responses

 | [hide all](#) | [show all](#)

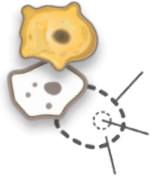
**RE: New Term Proposal: Exercise Study Facility** by whetzel 3 months ago

Can you expand on what an exercise study is?  
[reply](#)

**Exercise study facility definition** by Mette 3 months ago

A facility or core that provides services and equipment related to exercise physiology  
[reply](#)

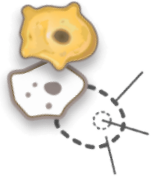
[Add Reply](#)



# Enhancing the Curation Workflow

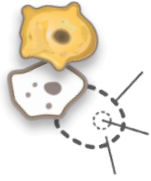
- Data submission
  - Ontology Widgets and Ontology Web services
- Ontology enrichment
  - BioPortal Notes
- Annotation of textual metadata
  - NCBO Annotator





# Tools for annotation of textual metadata

- NCBO Annotator
  - Open access, ontology-based Web service that annotates or “tags” textual metadata
  - Annotation is done using ontologies from BioPortal, which includes OBO Foundry and Unified Medical Language System ontologies
  - Variety of parameters that can be customized



# Annotator: The Basic Idea

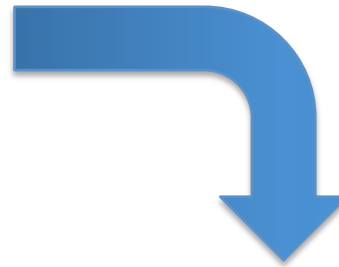
- Tag textual metadata with ontology terms

NCBI GEO > [Accession Display](#)

Scope:  Format:  Amount:  GEO accession:

Series GSE12391 [Query DataSets for GSE12391](#)

Status: Public on Sep 30, 2009  
Title: Melanoma: comparison between common nevi, radial/vertical growth phase melanoma, metastases and dysplastic nevi  
Organism: [Homo sapiens](#)  
Experiment type: Expression profiling by array  
Summary: **Despite malignant cutaneous melanoma's relative rarity compared to other skin cancers, it is still responsible for 80% of all skin cancer-related deaths. To identify molecular signatures of melanoma progression, excisional biopsies from 18 common melanocytic nevi (CMN), 8 primary radial growth phase melanomas (RGPM), 15 primary vertical growth phase melanomas (VGPM) and 5 melanoma metastases (MTS) were profiled using whole genome oligo-microarrays. Differentially expressed genes for each progression step were identified, and validation of selected transcripts by qRT-PCR was performed on an independent cohort of fixed samples. The comparison between CMN and RGPM showed an enrichment of Gene Ontology (GO) terms related to inter and intra-cellular junctions, whereas the transition from RGPM to VGPM was characterized by the deregulation of WNT3, MAPK and AKT pathways. In this step, enrichment analysis underlined the alteration of biological processes linked to apoptosis. Upregulation of genes involved into DNA double-strand breaks repair and downregulation of cellular adhesion genes were observed in**



**Melanoma Genes Malignant melanoma**

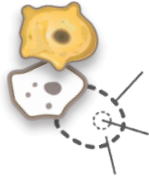
**growth & development Structural gene Growth Skin Skin Skin Neoplasms**

**Nevus Neoplasms Abnormal or excess nevi DNA Oncogenes Genome Biological Processes Genome**

**Risk Gene Expression Overall Heterogeneous analysis Intermediate Vertical Apoptosis Biopsy**

**Gene Expression Profiling DNA Mismatch Repair Neoplasm Metastasis Nevus, Pigmented Down-Regulation Up-Regulation**

**Deoxyribonucleic acid Pigmented nevi Circumpennate Polymerase Chain Reaction Dysplastic Nevus Syndrome Family**



```
1 use HTTP;
2 use URI::Escape;
3 use URI::Escape;
4 use XML::LibXML;
5
6 $|=1;
7
8 # The desired service URL
9 # http://rest.bioontology.org/oba/annotator
10 # http://stagerest.bioontology.org/oba/annotator
11
12 my $annotatorURL = 'http://stagerest.bioontology.org/oba/annotator';
13
14 # Sample text
15 my $text = uri_escape("Melanoma is a malignant tumor of melanocytes which are found predominantly in skin");
16
17 # create a user agent
18 my $ua = new LWP::UserAgent;
19
20 # create a parse to handle the output
21 my $parser = XML::LibXML->new();
22
23 # create a POST request
24 my $req = new HTTP::Request POST => "$annotatorURL";
25 $req->content_type('application/x-www-form-urlencoded');
26
27
28 # Set parameters
29 $req->content($paramsOnly=true"
30 "wholeWordOnly=true"
31 "score=true"
32 "ontology#toExpand=SNOMEDCT,NCI,MSH4"
33 "ontology#toKeepInResult=SNOMEDCT,NCI,MSH4"
34 "semanticType=DS1,DS4,DS194"
35 "selectOntologies=fast"
36 "watchDefaultStopWords=true"
37 "fontSize=14"
38 "levelMax=04"
39 "mappingType=ml14"
40 "mappingSource=genetod.edu");
41
42 # send request and get response.
43 my $res = $ua->request($req);
44
45 # Check the outcome of the response
46 if ($res->is_success) {
47     my $time = localtime();
48     print "Call successful at $time\n";
49 }
```

Code

### BioNLP UIMA Component Repository

BIONLP-UIMA HOME | DOWNLOAD | ABOUT US | BIONLP HOME |

#### BioNLP UIMA Component Repository

#### DOWNLOAD BIONLP-UIMA

The BioNLP Unstructured Information Management Architecture (UIMA) Component Repository provides UIMA wrappers for novel and well-known 3rd-party NLP tools used in biomedical text processing, such as tokenizers, parsers, named entity taggers, and tools for evaluation.

The following is a list of tools that have been integrated, with links to their websites:

**Analysis Engines:**

- Gene Identification: ABGene, ABNER, LingPipe, KeX
- Biomedical Term Recognition: NCBO Annotator Web Service
- Mutation Identification: MutationFinder
- Semantic Parsing: OpenDMAP
- Sentence Detection: KeX LingPipe OpenNL
- Tokenization: Genia Tagger, LingPipe, Penn

**Collection Readers** for corpora: Bio1 BioIE, Text

*The BioNLP-UIMA Component Repository code is licensed under the Apache License v2.0. Please note that the 3rd party res...*

UIMA platform

#### Annotator

Annotations: Annotate Text | Get Annotations By Resource Element Alpha

Ontologies: All Ontologies Selected | Choose...

Semantic Types: All Semantic Types Selected | Choose...

Options: Change...

The Annotator user interface is currently limited to 300 words. Please use the [NCBO Annotator web service](#) for more advanced options.

Text: Despite **melaniant cutaneous melanoma** is relatively rare compared to other skin cancers, it is still responsible for 80% of all skin cancer-related deaths. To identify **molecular signatures of melanoma progression**, excisional biopsies from 18 common melanocytic nevi (CMN), 8 primary radial growth phase melanomas (RGPM), 15 primary vertical growth phase melanomas (VGPM) and 5 melanoma metastases (MTS) were profiled using whole genome

Annotation Tag Cloud: (89)

- Adverse Event Ontology
- Ascomycete phenotype ontology
- Basic Formal Ontology
- Bilateria anatomy
- Biomedical Resource Ontology
- BioPortal Metadata
- BioTop

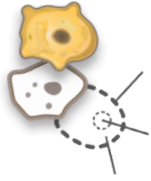
Annotation Tag Cloud: Melanoma Melan... Gene THE melanoma

User Interface

Term Id	Term Name	Score	Ontology Id	Ontology Name	Start	End	H
16 40403/407006005	Efalizumab	20	40403	SNOMED Clinical Terms	66	75	
17 40403/407005009	Efalizumab	20	40403	SNOMED Clinical Terms	253	262	
18 40403/407005009	Efalizumab	20	40403	SNOMED Clinical Terms	66	75	
19 40403/407006005	Efalizumab	20	40403	SNOMED Clinical Terms	253	262	
20 40403/9014002	Psoriasis	10	40403	SNOMED Clinical Terms	158	166	
21 40403/1522000	Plaque	10	40403	SNOMED Clinical Terms	151	156	
22 40403/260224007	Objective	10	40403	SNOMED Clinical Terms	5	13	
23 40403/410535002	Indicated	10	40403	SNOMED Clinical Terms	101	109	
24 40403/30207005	Risk of	10	40403	SNOMED Clinical Terms	58	64	
25 40403/200965009	Plaque psoriasis	10	40403	SNOMED Clinical Terms	151	166	

Excel

# Annotator Web service



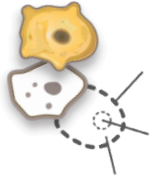
# Annotator Addin

The screenshot shows the Microsoft Excel interface with the Annotator Addin. The ribbon includes Home, Insert, Page Layout, Formulas, Data, Review, and View. The Annotator Service is active, showing a text input field and an 'Annotate' button. The spreadsheet data is as follows:

	A	B
1	16984639	The objective of this analysis was to assess the benefit:risk of efal...
2	20224509	A 49-year-old female was started on efalizumab for severe psoriasi...
3	19707339	The humanized antibody efalizumab is currently the only T-cell dire...
4	18550857	The kinase inhibitor dasatinib induces apoptosis in chronic lymphoc...
5	19016717	The present study demonstrated that dasatinib inhibited proliferat...
6	17962511	Herein, we show that dasatinib inhibits TCR-mediated signal transd...
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

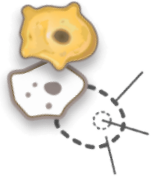
The Annotator Settings dialog box is open, showing the 'Ontologies' tab. It includes a 'Filter' field, a 'Choose Ontologies' list, and a 'Selected Ontologies' list. The 'Selected Ontologies' list contains 'SNOMED Clinical Terms [40403]'. The 'Choose Ontologies' list includes: Electrocardiography Ontology [4057], Tick gross anatomy [38366], SNP-Ontology [39215], African Traditional Medicine [40223], GeoSpecies Ontology [39933], MeGO [39976], Pseudogene [38873], C. elegans development [25148], Family Health History Ontology [3863], Ontology for Genetic Interval [40117], Sample processing and separation te..., MedlinePlus Health Topics [40397], BRENDA tissue / enzyme source [40..., and Minimal anatomical terminology [402...]. The dialog box also has 'OK' and 'Cancel' buttons.





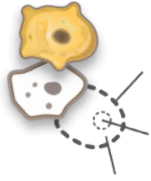
# Enhancing the Curation Workflow

- Data submission
  - Ontology Widgets and Ontology Web services
- Ontology enrichment
  - BioPortal Notes
- Annotation of textual metadata
  - NCBO Annotator



# Acknowledgements

- NCBO Team
  - Mark Musen, Stanford University
  - Partners: Barry Smith, University of Buffalo, Chris Chute, Mayo Clinic, and Peggy Storey, University of Victoria
  - Developers, Driving Biological Projects, and other Collaborators



# Thank you!

- Using NCBO Technology in Your Project:
  - [http://www.bioontology.org/wiki/index.php/Using\\_NCBO\\_Technology\\_In\\_Your\\_Project](http://www.bioontology.org/wiki/index.php/Using_NCBO_Technology_In_Your_Project)
- Web service documentation:
  - [http://www.bioontology.org/wiki/index.php/NCBO\\_REST\\_services](http://www.bioontology.org/wiki/index.php/NCBO_REST_services)
- Questions:
  - [support@bioontology.org](mailto:support@bioontology.org)



