

A Methodology to Migrate the Gene Ontology to a Description Logic Environment using DAML+OIL

**Chris Wroe, Robert Stevens,
Carole Goble**

University of Manchester, UK

Michael Ashburner

EBI, Hinxton, UK

<http://gong.man.ac.uk>

Gene Ontology Next Generation Project (GONG)

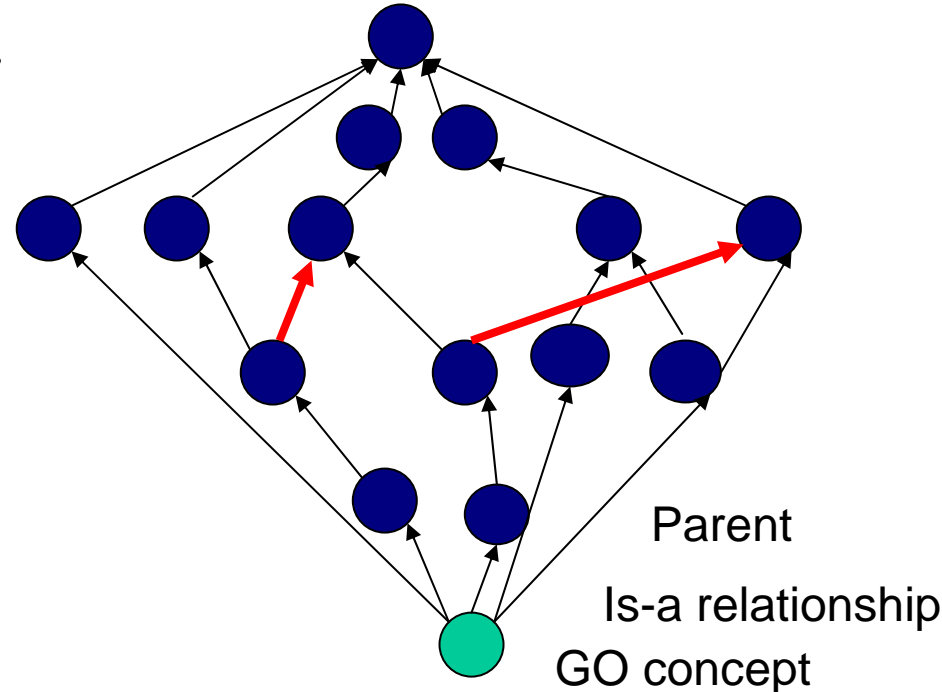
- Demonstrate the utility of finer grained concept descriptions in DAML+OIL
- Develop methodologies and tools to support the process

Translating theory into practice

- Gene Ontology provides a service to the model organism database community
- Description logic (DL) is a technology born out of computer science research
- DAML+OIL becoming a standard ontology interchange language underpinned by DL
- Adopted by W3C and will soon become Ontology **W**eb **L**anguage (**OWL**)

GONG - proof of concept

- Maintaining an exhaustive is-a structure



Example: heparin biosynthesis

[*chemical*] biosynthesis (GO:0009058)

Axis 1:

[i] *carbohydrate* biosynthesis (GO:0016051)

Chemicals

[i] *aminoglycan* biosynthesis (GO:0006023)

[i] *heparin* biosynthesis (GO:0030210)

Example: heparin biosynthesis

[*chemical*] biosynthesis (GO:0009058)

Axis 1:

[i] *carbohydrate* biosynthesis (GO:0016051)

Chemicals

[i] *aminoglycan* biosynthesis (GO:0006023)

.....

[i] *heparin* biosynthesis (GO:0030210)

Axis 2:

Process

[i] *heparin metabolism* (GO:0030202)

[i] *heparin biosynthesis* (GO:0030210)

Example: heparin biosynthesis

[*chemical*] biosynthesis (GO:0009058)

Axis 1:

[i] *carbohydrate* biosynthesis (GO:0016051)

Chemicals

[i] *aminoglycan* biosynthesis (GO:0006023)

[i] *glycosaminoglycan* biosynthesis (GO:0006024)

[i] *heparin* biosynthesis (GO:0030210)

Axis 2:

Process

[i] *heparin metabolism* (GO:0030202)

[i] *heparin biosynthesis* (GO:0030210)

<http://gong.man.ac.uk>

Is this important?

- Missing is-a not noticed by users
- BUT... improves fidelity of DB record retrieval.
 - Asking for gene products involved in 'glycosaminoglycan biosynthesis' will lead to an additional result:

O94923 SPTTr ISS - D-glucuronyl C5-epimerase (Fragment)

How can DAML+OIL support the task?

- Step 0. Translate to DAML+OIL syntax
 - Provided by OilEd (ontology editing tool)

<http://oiled.man.ac.uk>

GO RDF	DAML+OIL
<code><go:term></code>	<code><daml:Class></code>
<code><go:isa></code>	<code><daml:subClassOf><daml:Class></code>
<code><go:part-of></code>	<code><daml:subClassOf><daml:Restriction></code> <code> <daml:onProperty><daml:ObjectProperty</code> <code> rdf:resource="go:part-of"/></code> <code> <daml:hasClass><daml:Class></code>

<http://gong.man.ac.uk>



THE UNIVERSITY
of MANCHESTER

DAML+OIL definitions for metabolism concepts

- heparin biosynthesis

- **class** heparin biosynthesis **defined**
subClassOf biosynthesis
restriction onProperty acts_on **hasClass** heparin
(acts_on is unique)
- Paraphrase: biosynthesis which acts solely on heparin

- glycosaminoglycan biosynthesis

- class glycosaminoglycan biosynthesis defined
subClassOf biosynthesis
restriction onProperty acts_on hasClass **glycosaminoglycan**

A chemical ontology

- Mapped chemical concepts to MeSH (using UMLS tools/ API e.g. Norm).
- Created a DAML+OIL ontology from a subset of the MeSH chemical taxonomy
- Provides the following information:

carbohydrates

[i] polysaccharides

[i] glycosaminoglycans

[i] heparin

<http://gong.man.ac.uk>



THE UNIVERSITY
of MANCHESTER

Paraphrased reasoning process

- heparin biosynthesis

- class heparin biosynthesis defined
subClassOf biosynthesis
restriction onProperty acts_on hasClass **heparin**

- glycosaminoglycan biosynthesis

- class glycosaminoglycan biosynthesis defined
subClassOf biosynthesis
restriction onProperty acts_on hasClass **glycosaminoglycan**

Is-a



Inferring a new is-a link

- **heparin biosynthesis**

- class heparin biosynthesis defined
subClassOf biosynthesis
restriction onProperty acts_on hasClass heparin

Is-a

- **glycosaminoglycan biosynthesis**

- class glycosaminoglycan biosynthesis defined
subClassOf biosynthesis
restriction onProperty acts_on hasClass glycosaminoglycan

Is-a

Output

- OilEd API reports additional inferred is-a relationships.

E.g.

heparin biosynthesis has new is-a parent
glycosaminoglycan biosynthesis

- Report sent to GO editorial team for comment.
- They makes changes to GO if appropriate and sends back queries

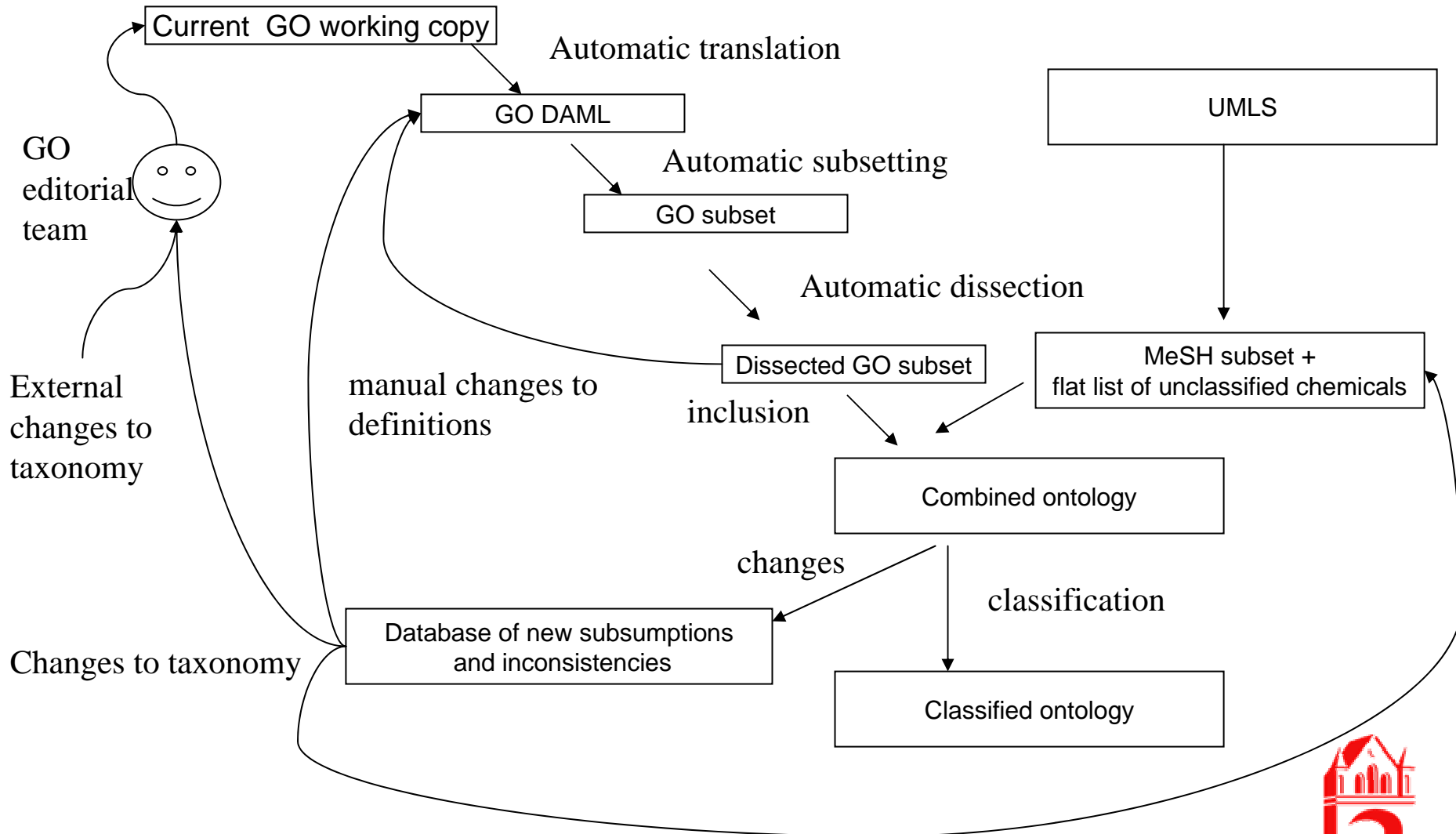
Results

- Carbohydrate metabolism ~250 concepts
 - 22 additional is-a links 17 of which now in GO
- Amino acid metabolism ~ 250 concepts
 - Further 17 additional is-a links now in GO
- GO team will be reviewing results for metabolism as a whole once we have the tools to support the process
- Useful results come from even a partial coverage

Build a practical environment

- Tools needed for:
 - Creating DAML+OIL definitions
 - Tracking changes
 - Reporting reasoning results
 - Viewing definitions

Multi-step workflow



<http://gong.man.ac.uk>



THE UNIVERSITY
of MANCHESTER

Reporting tools

The image displays three screenshots of a web browser showing an ontology interface.

Left Window: Shows a hierarchical list of metabolic processes. The address bar is `http://localhost:8080/jyserv/ontology.py?contract=proteoglycan%20metabolism#proteoglycan%20metabolism`. The list includes:

- carbohydrate metabolism
 - + polysaccharide metabolism
 - + disaccharide metabolism [classification change](#)
 - + proteoglycan metabolism [classification change](#)
 - + oligosaccharide metabolism [classification change](#)
 - + glucuronoside metabolism
- aminoglycan metabolism
 - + aminoglycan biosynthesis
- glycosaminoglycan metabolism [classification](#)
 - + hyaluronan metabolism
 - + chondroitin sulfate metabolism
 - + dermatan sulfate metabolism
 - glycosaminoglycan catabolism [classification](#)
- heparin metabolism
 - heparin biosynthesis [classification change](#)
 - heparin catabolism [classification change](#)
- glycosaminoglycan biosynthesis [classification](#)
 - keratan sulfate biosynthesis
- + aminoglycan catabolism
- + organic acid metabolism
- + peptidoglycan metabolism [classification change](#)
- + alditol metabolism [classification change](#)
- + aldonic acid metabolism
- + aminosugar metabolism
- mannan metabolism [classification change](#)

Middle Window: Shows a search for 'heparin biosynthesis'. The address bar is `http://localhost:8080/jyserv/ontology.py?clsfychange=glycosaminoglycan%20biosynthesis`.

Right Window: Shows the results for 'heparin biosynthesis'. The address bar is `http://localhost:8080/jyserv/ontology.py?clsfychange=glycosaminoglycan%20biosynthesis`. The results include:

Original parents:

- heparin metabolism

Post classification parents:

- heparin metabolism
- glycosaminoglycan biosynthesis

heparin biosynthesis

Created using the pattern: pattern (.+?) ([a-z]+ism\$|[a-z]+ing\$|[a-z]+tion\$|[a-z]+sis\$|[a-z]+age\$)

From source: ontology URI
file:/d:/documents/gong/metabolism_ontologies/jul_02_metab...
class URI <http://www.geneontology.org/go#heparin>
biosynthesis; ontology version 0;

RUBIRC "heparin biosynthesis"
CODE "GO:0030210"
DOC "Biosynthesis"
IS-A [heparin metabolism](#) ; GO:0030202 ; process
IS-A [biosynthesis](#) ; GO:0009058 ; process
defining ACTS_ON [heparin](#) ; C0019134 ; chemical

<http://gong.man.ac.uk>

Using Amigo to view definitions



The screenshot shows the AmiGO web interface. At the top, there is a search bar labeled "Search GO:" and a button labeled "Ex". Below the search bar, there are links for "Top Docs", "Gene Ontology", and "GO Links". The main content area displays a hierarchical tree structure for the Gene Ontology term **GO:0003673 : Gene Ontology (0)**. The tree is expanded to show several child terms, including **GO:0008150 : biological process (0)**, **GO:0008151 : cell growth and/or maintenance (0)**, **GO:0008152 : metabolism (0)**, **GO:0005975 : carbohydrate metabolism (0)**, **GO:0006022 : aminoglycan metabolism (0)**, **GO:0030203 : glycosaminoglycan metabolism (0)**, **GO:0030202 : heparin metabolism (0)**, **CO:00000613 : heparin (0)**, **GO:0030210 : heparin biosynthesis (0)**, **GO:0030211 : heparin catabolism (0)**, **GO:0005575 : cellular component (0)**, and **GO:0003674 : molecular function (0)**. At the bottom of the interface, there are links to "Get this tree as RDF XML.", "Get this data as a GO flat file.", and "Get a bookmarkable url of this tree."

AmiGO

Search GO:

[Top Docs](#) [Gene Ontology](#) [GO Links](#)

[-] **GO:0003673 : Gene Ontology (0)**

- [+] **GO:0008150 : biological process (0)**
- [+] **GO:0008151 : cell growth and/or maintenance (0)**
- [+] **GO:0008152 : metabolism (0)**
 - [+] **GO:0005975 : carbohydrate metabolism (0)**
 - [+] **GO:0006022 : aminoglycan metabolism (0)**
 - [+] **GO:0030203 : glycosaminoglycan metabolism (0)**
 - [+] **GO:0030202 : heparin metabolism (0)**
 - **CO:00000613 : heparin (0)**
 - [+] **GO:0030210 : heparin biosynthesis (0)**
 - [+] **GO:0030211 : heparin catabolism (0)**
- [+] **GO:0005575 : cellular component (0)**
- [+] **GO:0003674 : molecular function (0)**

[Get this tree as RDF XML.](#)
[Get this data as a GO flat file.](#)
[Get a bookmarkable url of this tree.](#)

<http://gong.man.ac.uk>



THE UNIVERSITY
of MANCHESTER

Conclusion & future plans

- Description logic approach is useful
- Beginning to build a practical environment
- Extend quantity and detail of DAML+OIL definitions
 - ~2000 complete but un-validated metabolism definitions
 - ~2000 partial enzyme definitions
- Continue prototyping and developing tools to support the process.

Acknowledgments

- Jane Lomax and Midori Harris of the GO editorial team for help and advice and responding to the suggested changes
- UMLS and MeSH which provided valuable resources for chemical information
- Sean Bechhofer for development on OilEd
- Project funded as a subcontract of the DARPA DAML programme