



# GOing forward: Refining Gene Ontology Content

Midori A. Harris  
EBI & GO Consortium  
Genome Informatics 2004

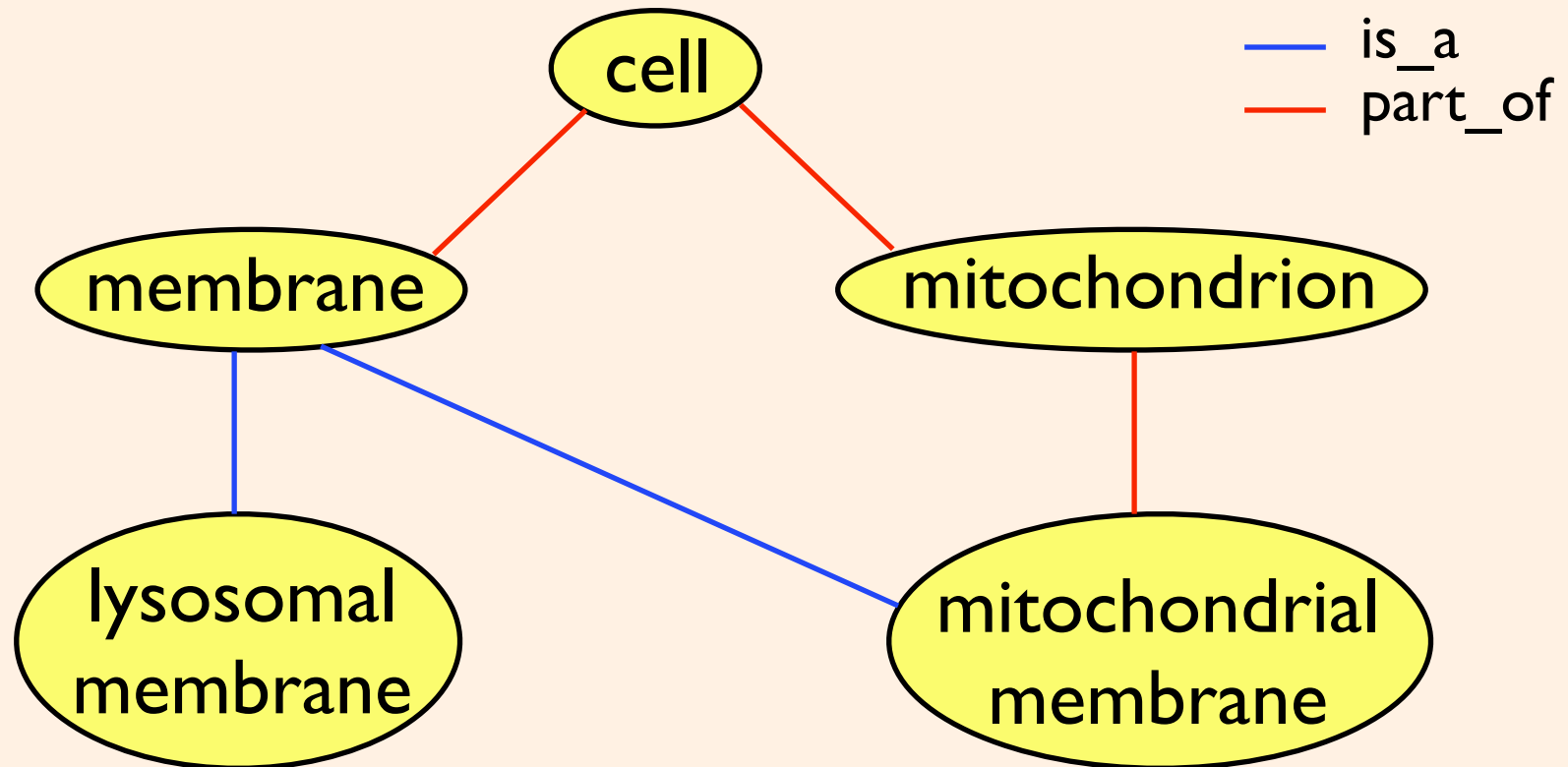


# What is GO?

- Species-neutral controlled vocabulary
- Three structured networks of defined terms to describe gene product attributes
  - molecular function
  - biological process
  - cellular component
- Used for gene product annotation



# GO Structure: DAG



GO is a directed acyclic graph (DAG):  
a term can have one or more parent(s)



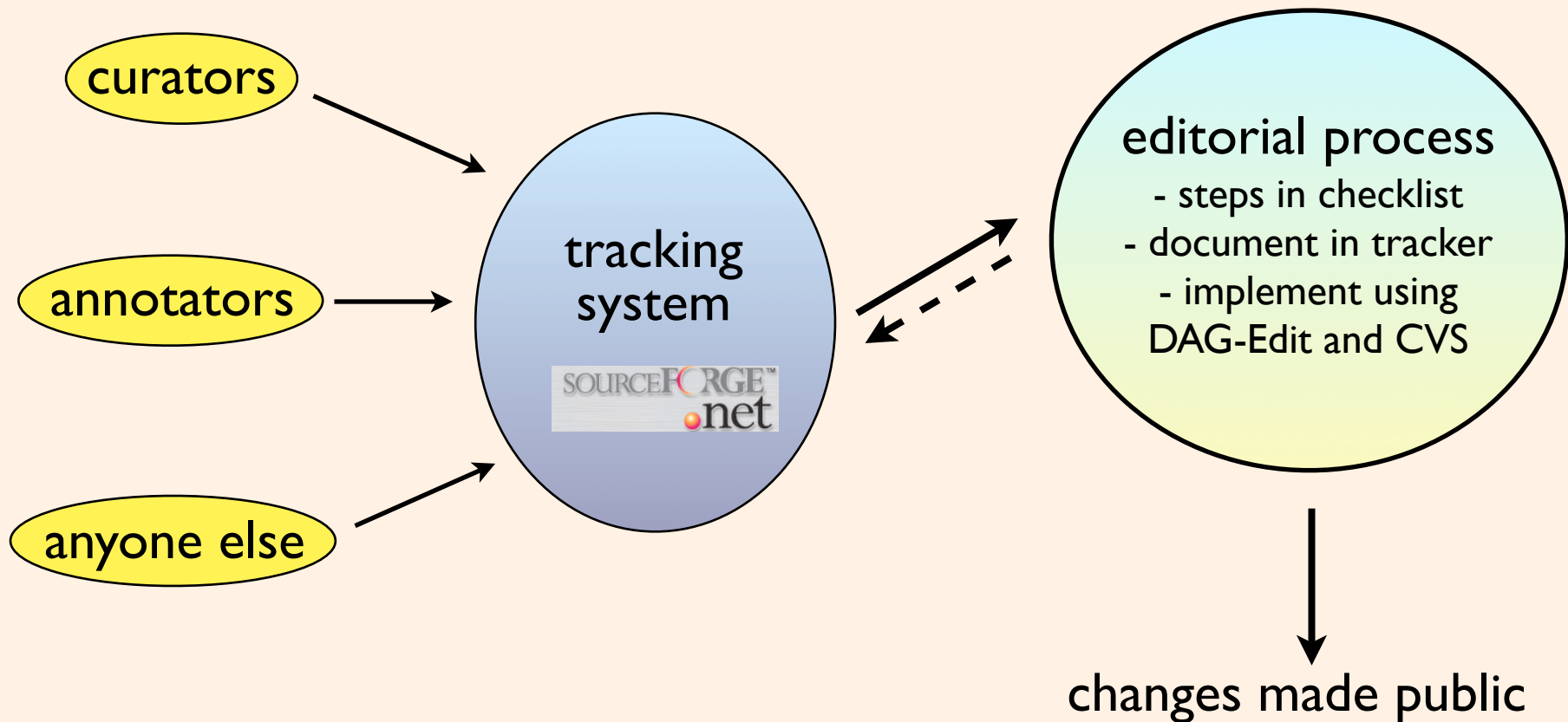
# Why GO Changes

- Advances in biology
- New groups join, requiring new terms or different relationships between terms
- Update legacy terms
- Improve logical consistency



# How GO Changes:

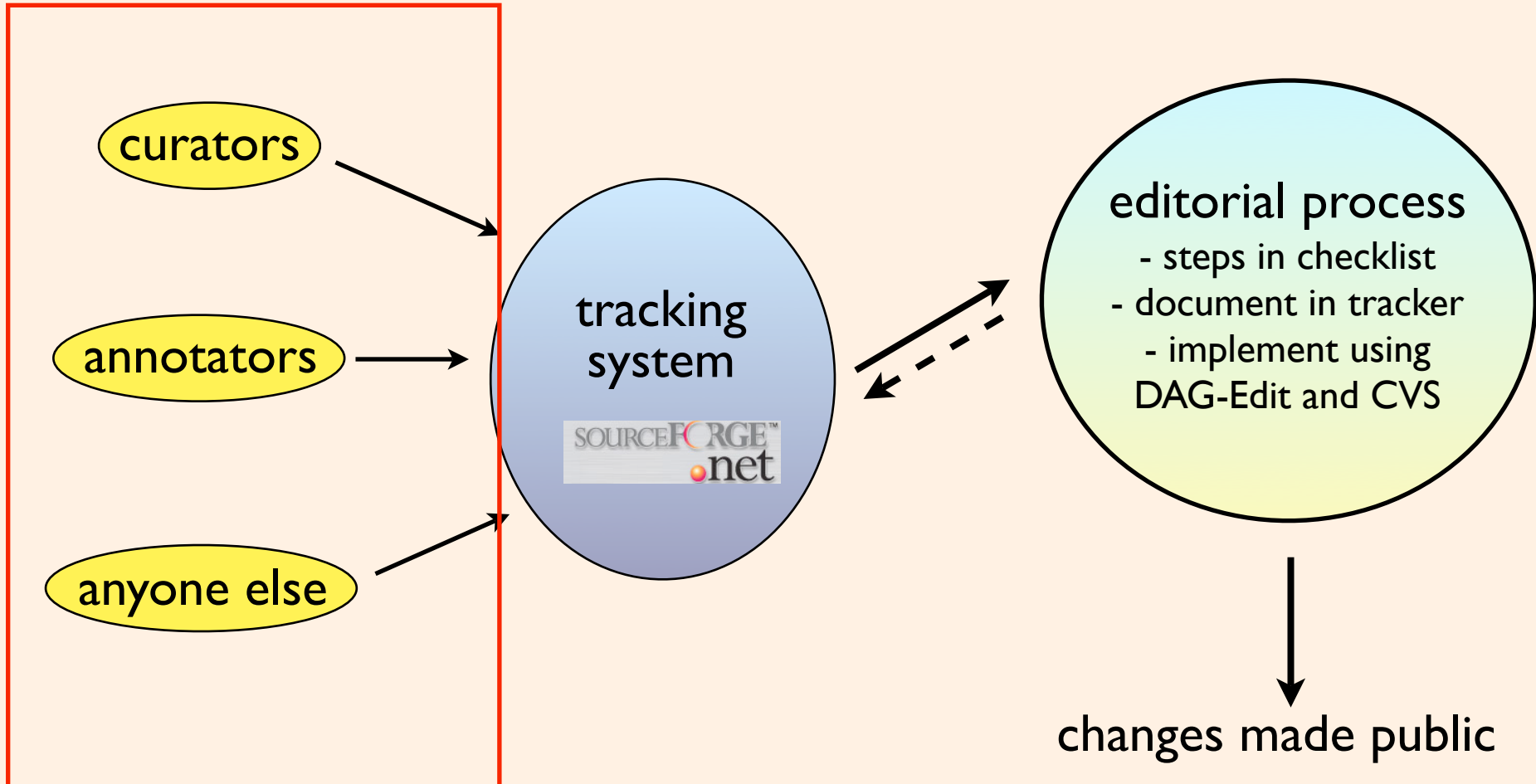
## Overview of the GO Editorial Procedure





# How GO Changes:

## Overview of the GO Editorial Procedure





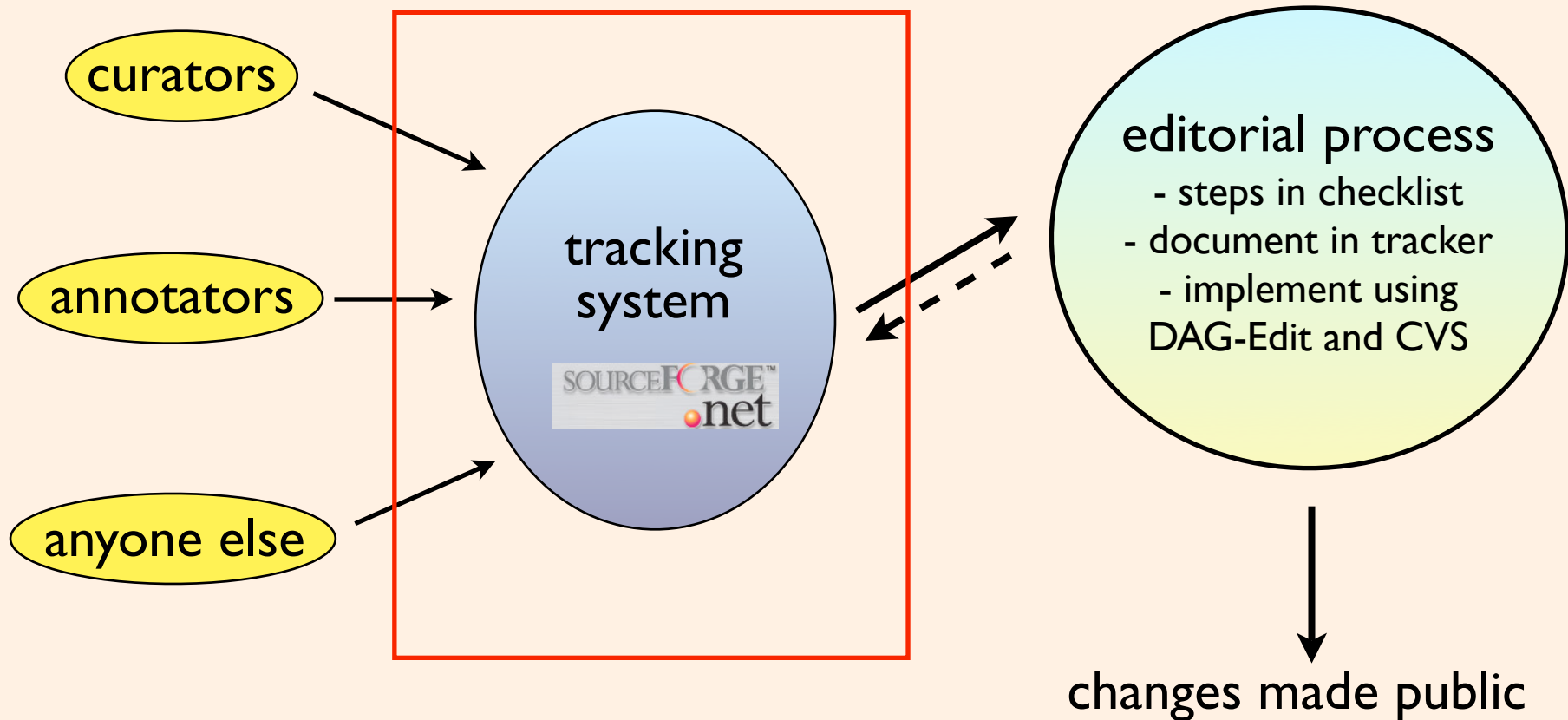
# Sources of Ideas

- Curators – implement decisions from meetings; ongoing maintenance; specific areas of interest
- Annotators – using GO to describe gene products
- Interest groups – involving experts in their areas of expertise
- Computation – parsing, reasoning



# How GO Changes:

## Overview of the GO Editorial Procedure







# SourceForge Tracker

- Item for each proposed change
- Submissions from curators, annotators, others
- Each item assigned to a curator
- Submitters and others can comment on items
- Record of discussion and progress
- Archive of all entries

[https://sourceforge.net/tracker/?func=add&group\\_id=36855&atid=440764](https://sourceforge.net/tracker/?func=add&group_id=36855&atid=440764)



# Curator Requests List (SourceForge)

SourceForge.net

https://sourceforge.net/tracker/?func=browse&group\_id=36855&atid=440764

OSTG | ThinkGeek - Slashdot - ITMJ - Linux.com - NewsForge - freshmeat - Newsletters -

SOURCEFORGE.net

my sf.net | software map | donate to sf.net | about sf.net | My Favorites | Go

welcomes gomidori (Logout)

**Register New Project**

**Search**

Curator requests | Search

powered by **YAHOO!** search

**SF.net Subscription**

- Subscribe Now
- Manage Subscription
- Advanced Search
- Direct Download
- Priority Tech Support
- Project Monitoring

**SF.net Resources**

- Site Docs
- Site Status (09/17)
- SF.net Supporters

**Project: Gene Ontology: Browse Curator requests**

Summary | Admin | Home Page | Forums | Tracker | Bugs | Support | Patches | RFE | Lists | Tasks | Docs | News | CVS | Files |

**Submit New | Browse | Reporting | Admin | Search**

Assignee: Any | Status: Open | Category: Any | Group: Any

Show only: Submitter username (show mine): | Summary keyword: |

Sort By: ID | Descending | Browse

Browse the term requests currently under consideration.

Request ID	Summary	Open Date	Priority	Assigned To	Submitted By
<input type="checkbox"/> 1031159	<a href="#">Response and Detection Terms</a>	2004-09-20 11:35	7	nobody	aledie
<input type="checkbox"/> 1029968	<a href="#">add parent?: cytokinesis, formation of actomyosin</a>	2004-09-17	5	nobody	val wood
<input type="checkbox"/> 1029917	<a href="#">nucleotide binding</a>	2004-09-17 15:11	5	nobody	jenclark
<input type="checkbox"/> 1029875	<a href="#">new term request: ciliary rootlet</a>	2004-09-17 13:39	5	beckyfoulger	beckyfoulger
<input type="checkbox"/> 1029477	<a href="#">potassium ion-transporting ATPase complex</a>	2004-09-16 20:36	5	nobody	mlgwin

**Site Sponsors**

[https://sourceforge.net/tracker/?func=add&group\\_id=36855&atid=440764](https://sourceforge.net/tracker/?func=add&group_id=36855&atid=440764)



# Curator Request Item (SourceForge)

SourceForge.net: Modify:1024533 - exocyst assembly

https://sourceforge.net/tracker/index.php?func=detail&aid=1024533&group\_id=36855&at=

OSTG | ThinkGeek - Slashdot - ITMJ - - Linux.com - NewsForge - freshmeat - Newsletters - - Advertisement

SOURCEFORGE™ net

welcomes gomidori (Logout)

[Register New Project](#)

**Search**

Curator requests

powered by **YAHOO!** search

**SF.net Subscription**

- [Subscribe Now](#)
- [Manage Subscription](#)
- [Advanced Search](#)
- [Direct Download](#)
- [Priority Tech Support](#)
- [Project Monitoring](#)

**SF.net Resources**

- [Site Docs](#)
- [Site Status \(09/17\)](#)
- [SF.net Supporters](#)
- [Compile Farm](#)
- [Foundries](#)
- [Project Help Wanted](#)
- [New Releases](#)
- [Get Support](#)

**Site Sponsors**

**Project: Curator requests: Trackers**

[Summary](#) | [Admin](#) | [Home Page](#) | [Forums](#) | [Tracker](#) | [Bugs](#) | [Support](#) | [Patches](#) | [RFE](#) | [Lists](#) | [Tasks](#) | [Docs](#) | [News](#) | [CVS](#) | [Files](#)

[Submit New](#) | [Browse](#) | [Reporting](#) | [Admin](#) | [Search](#)

[ 1024533 ] exocyst assembly

[Monitor](#)

**Submitted By:**  
Harold J. Drabkin - [hjdabkin](#)

**Last Updated By:**  
hjdabkin - Comment added

**Number of Comments:**  
4

**Data Type:**  
Curator requests

**Category:**  
New term request (admin)

**Assigned To:**  
None (admin)

**Status:**  
Open

**Summary:**  
exocyst assembly

I'd like to add a term describing the assembly of the exocyst, a protein complex peripherally associated with the plasma membrane that determines where secretory vesicles dock and fuse (GO: 0000145). First, however, I'd like to suggest that the definition of

**Date Submitted:**  
2004-09-08 17:03

**Date Last Updated:**  
2004-09-14 14:37

**Number of Attachments:**  
0

[Submit Changes](#)

**Group:**  
MGD (admin)

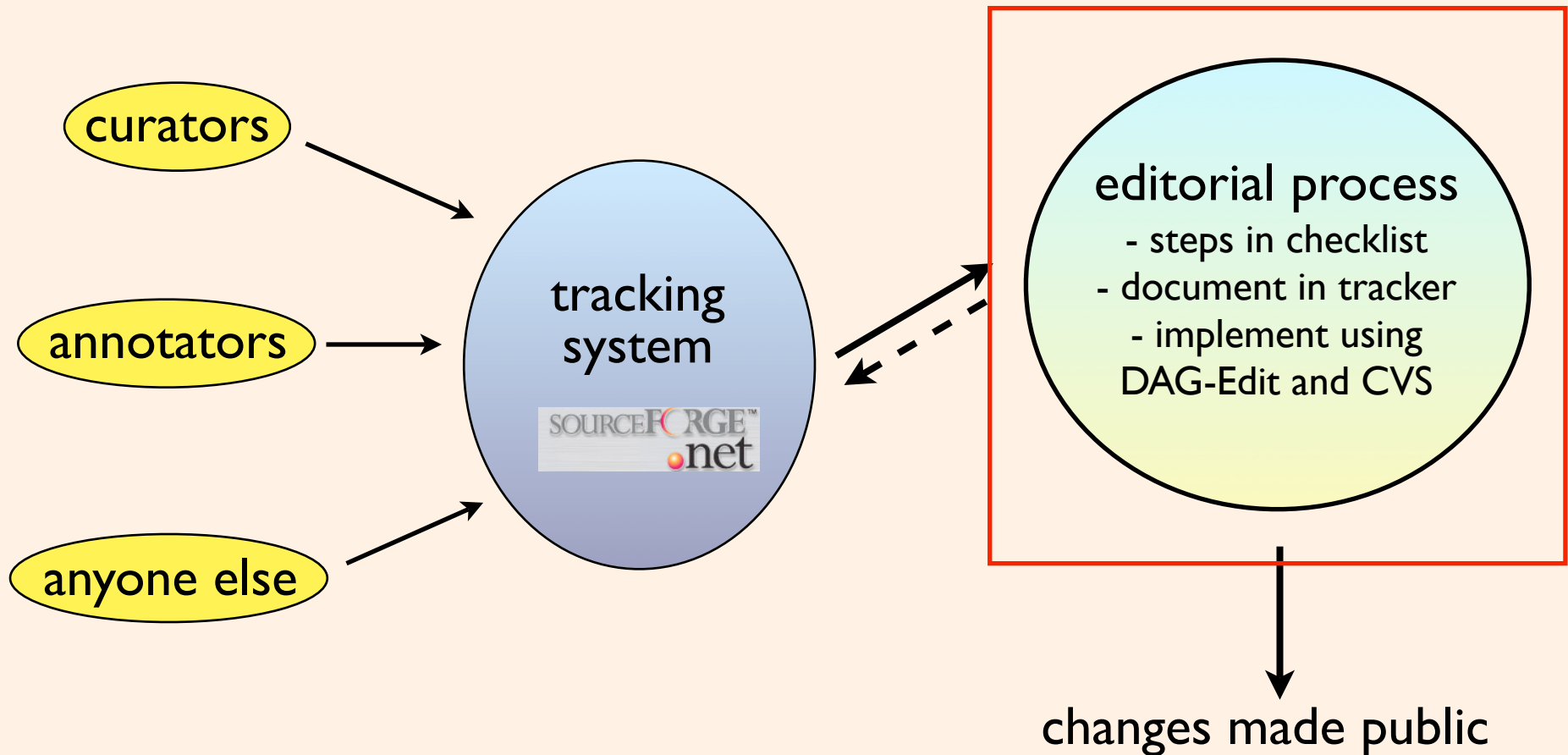
**Priority:**  
5 - Medium

**Resolution:**  
None



# How GO Changes:

## Overview of the GO Editorial Procedure





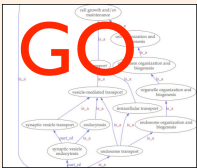
# GO Curation Checklist

- Claim a tracker item
  - Does the term belong in GO?
  - Term name & definition
    - standard wording?
  - Relationships to other terms
- Consult literature, interest group, other curators & researchers as needed



# GO Interest groups

- Cover specific topics within GO
- Include GO curators and outside experts
- Group members develop portions of the ontology, then report to GO Consortium
- Sample topics:
  - metabolism
  - developmental biology
  - plant biology
  - transcription



# GO Interest Groups

GO Curator Interest Groups

http://www.geneontology.org/GO.interests.html

## GO Curator Interest Groups

Autophagy | Cell Cycle | Cell Growth and/or Maintenance | Conjugation | Cytoskeleton and Cell Motility | Developmental Biology | DNA Repair | Enzymes | Fungal & Microbial Cell Type & Structures | GPCRs | Hemopoiesis | Immunology | Meiosis | Metabolism | Mitochondrion | Pathogen | Physiology | Plants | Plant-associated microbes | Protein Folding | Protein Kinases | Protein Modification | Protein Synthesis | RNA Metabolism | Signal Transduction | Stress Res

GO

OBO  
open biological ontologies

[Open all menus](#)  
[Site map](#)  
[Home](#) | [New](#) | [FAQ](#)

[Downloads](#)  
.. [Current Ontologies](#)  
.. [Current Annotations](#)  
.. [GO Database](#)

[Documentation](#)  
[GO Tools](#)  
[Mappings to GO](#)

[Archives](#)

[About GO](#)  
.. [GO Consortium](#)  
.. [Mailing lists](#)  
.. [Publications](#)  
.. [GO People](#)  
.. [Interest Groups](#)  
.. [Future Meetings](#)  
.. [Career Opportunities](#)  
.. [Contact Details](#)  
.. [Funding](#)  
.. [Acknowledgements](#)  
[Terms of Use](#)  
[Contact GO](#)  
[Report Errors](#)

Search  
Terms/Annotations

Search Site

GO Curator Interest Groups

http://www.geneontology.org/GO.interests.html

## Introduction

At the GO meeting in Cambridge, Sept. 2002, we decided to divide up w this arrangement, we'll set up "interest groups" for any areas within the or or revisions, or to have proposed changes crop up frequently. Proposed c handled (or at least seen) by that group.

Membership is open to anyone, and all members are listed on the web sit [list](#). For complete archiving, discussions may be carried out on the GO lis should begin with the name of the interest group, followed by a hyphen. I

1) Plants interest group's discussion of development and gametogenesis:

**Subject: Plants - Development and gametogenesis**

2) Transcription interest group's discussion of the definition of transcripion

**Subject: Transcription - definition**

A mail window with the appropriate subject line will be generated autom group that you wish to address. If discussions occur outside of the GO ma list for archiving.

To set up a new interest group please mail [Midori Harris](#).

*Note: These interest groups are organized by topics that roughly corres appear in GO Slim sets. Organism-specific ("sensu") terms are not listec organism database affiliations, but "sensu" terms should be treated in an relevant MOD curators should be consulted.*

## Current Interest Groups and Members:

### [Autophagy \(process\)](#)

### [Plant-associated microbes](#)

Covers biological functions, processes and structures that microbes use to form associat between this group and the [Pathogens](#) and [Fungal and microbial cell](#) groups.

[David Bird](#) (North Carolina State University)  
[Robin Buell](#) (TIGR)  
[Allan Collmer](#) (Cornell University)  
[Candace Collmer](#) (Wells College)  
[Ralph Dean](#) (North Carolina State University)  
[Michelle Gwinn](#) (TIGR)  
[Linda Hannick](#) (TIGR)  
[Nicole Perna](#) (University of Wisconsin)  
[Joao Setubal](#) (Virginia Bioinformatics Institute)  
[Brett Tyler](#) (Virginia Bioinformatics Institute)  
[Emily dimmer](#) (GOA)  
[Jennifer Clark](#) (EBI GO)

### [Protein folding \(process\)](#)

[Rama Balakrishnan](#) (SGD)  
[Linda Hannick](#) (TIGR)  
[Val Wood](#) (Sanger)

### [Protein kinases \(function\)](#)

[Rama Balakrishnan](#) (SGD)  
[Becky Foulger](#) (FlyBase)  
[Sue Rhee](#) (TAIR)



# DAG-Edit

The screenshot displays the DAG-Edit version 1.419-beta3 interface. The window title is "DAG-Edit version 1.419-beta3". The menu bar includes "File", "Edit", "Plugins", and "Help".

**tree view:** A hierarchical tree on the left side shows biological processes and cellular components. The "cellular\_component" folder is expanded, showing "cell" and "mitochondrial membrane". The "mitochondrial membrane" node is selected and highlighted in blue.

**search:** A search bar at the top right contains the text "search". Below it, the "Find terms" section shows a search query: "AND [0] Term has self any id that ends with 004659". There are checkboxes for "Advanced", "All", "Selected", "Children", and "Obsoletes".

**editing:** The central panel shows the details for the selected term "mitochondrial membrane" (ID: GO:0005740). It includes a "Definition" tab with the text "The lipid bilayer surrounding the mitochondrion and separating its" and a "Comment" field with "GO:ai". There are "Add" and "Del" buttons for synonyms and general DbXrefs.

**DAG view:** A "DAG Viewer" panel on the right shows a hierarchical view of the selected term's relationships. It shows "cellular\_component" containing "cell", which contains "intracellular", "cytoplasm", "mitochondrion", and "membrane". "mitochondrial membrane" is highlighted in blue.

Other visible elements include a "GO Obsolescence Plugin v1.003" at the bottom left with a "Cannot obsolete nodes with children" message, and a "Dbxref Library v1.002" at the bottom right with fields for "GO:mah", "ISBN:0", and "PMID:0".





# Tracker Statistics

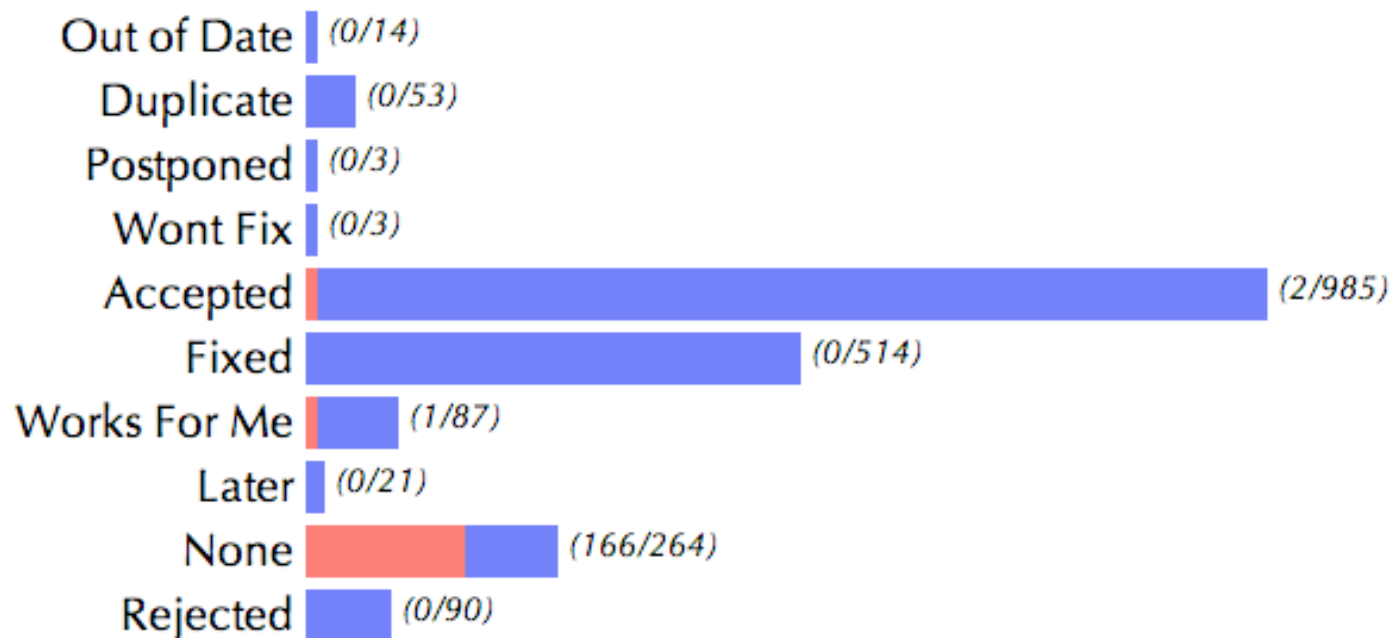
- September 23, 2004
  - 164 Curator Request items open
  - 2038 Total
- Reporting available online:

[https://sourceforge.net/tracker/reporting/?group\\_id=36855&atid=440764](https://sourceforge.net/tracker/reporting/?group_id=36855&atid=440764)



# Request Resolution

## Distribution By Resolution



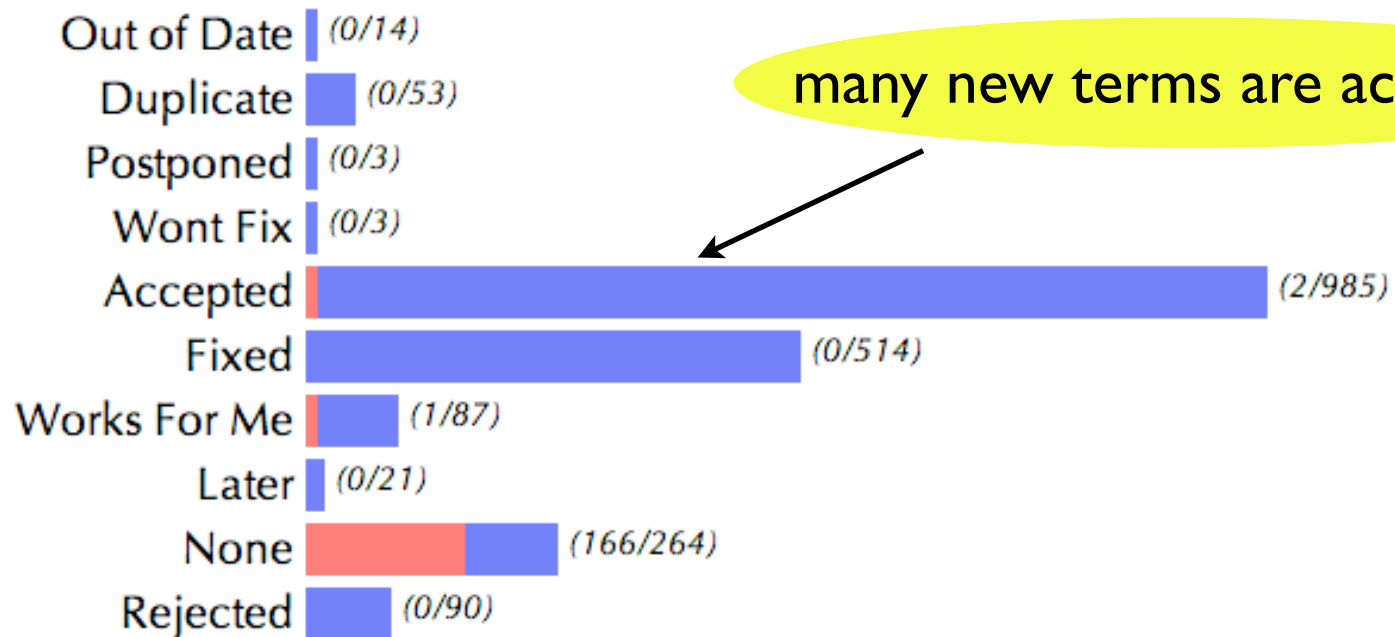
Key: ( ■ Open / ■ All )

Sept. 22, 2004



# Request Resolution

## Distribution By Resolution



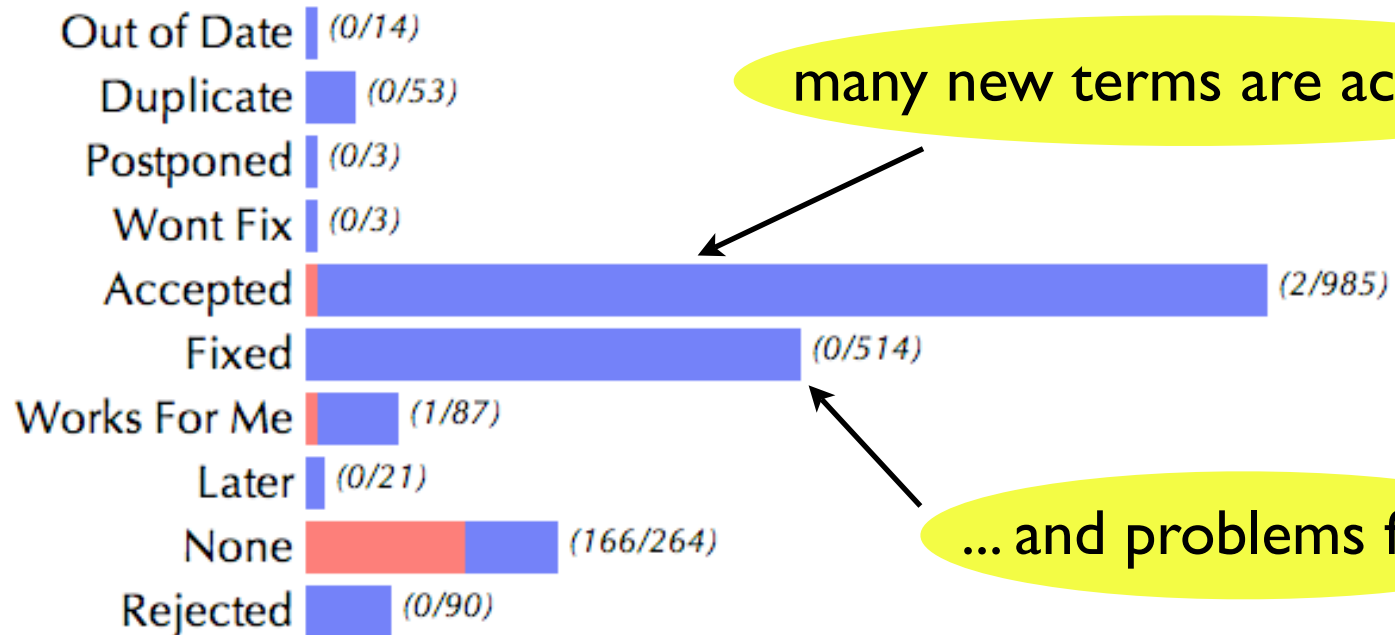
Key: ( ■ Open / ■ All )

Sept. 22, 2004



# Request Resolution

## Distribution By Resolution



many new terms are accepted

... and problems fixed

Key: ( ■ Open / ■ All )

Sept. 22, 2004



# Current Priorities I

- Biology topics
  - cell cycle
  - cell killing and pathogenesis
  - metabolism: cellular vs. organismal
  - developmental biology
  - cellular component – top-level terms



# Current Priorities 2

- Other content changes
  - new relationship types
  - is\_a parents for all terms
  - corrections to part\_of relationships
  - results of OBOL parsing and other computational approaches
  - missing relationships



# Using OBOL

- OBO Language
  - term decomposition using grammars
  - generating computable logical class definitions
  - rules and reasoning over class definitions
- Builds on existing structure in GO terms
- Can be used to complement manual ontology maintenance
  - find missing relationships
  - correct errors
  - several possible ways to implement

*from Chris Mungall, BDGP*



# Acknowledgements

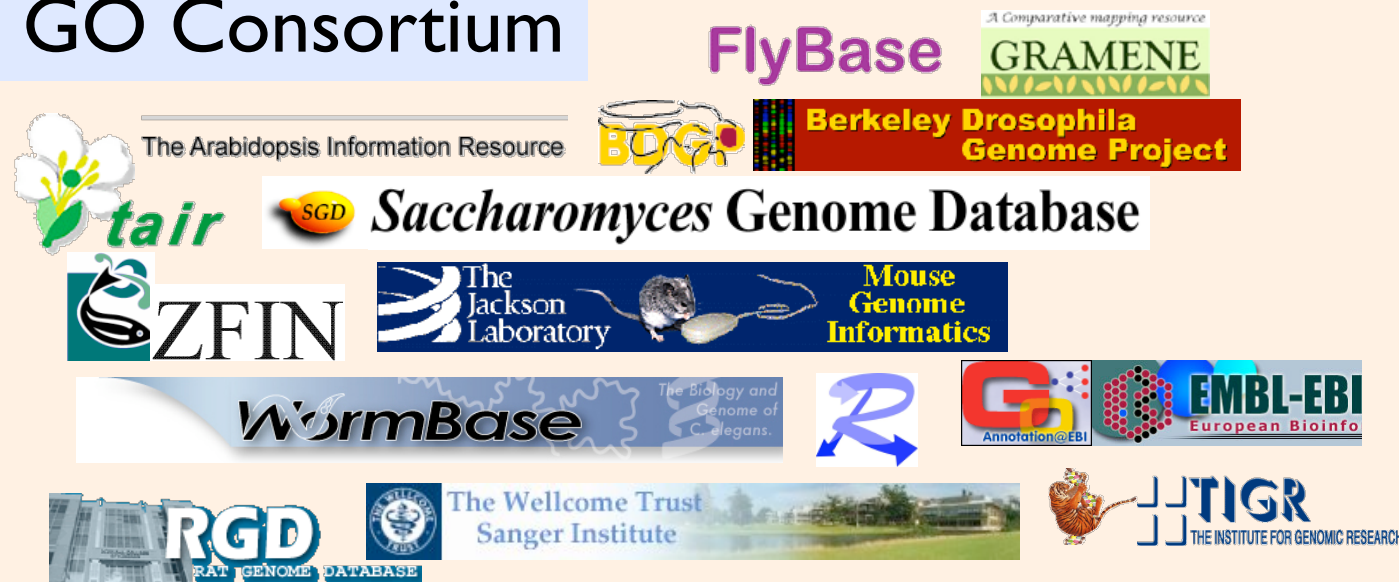
The GO Editorial Office



Jane Lomax, Amelia Ireland, Jennifer Clark

---

The GO Consortium



[www.geneontology.org](http://www.geneontology.org)

The Gene Ontology Consortium receives funding from the National Human Genome Research Institute, AstraZeneca, Incyte Genomics, the European Union and the UK Medical Research Council.





# Curation Results I

- “Easy” items:
  - Unambiguous request
  - Clear term name(s) and meanings
  - Clear relationships to other terms
  - Curators make change(s) at once
- *Example: cobalt ion binding (SourceForge entry 896544)*



# Curation Results 2

- Moderately difficult items:
  - Some aspect of request requires clarification
  - Different proposals to be resolved
  - Discussion via SourceForge comments and email before curators make change(s)
- *Example: fore-, mid-, and hindbrain development (SourceForge entry 854736)*



# Curation Results 3

- Very difficult items
  - Challenging biology to model
  - Extensive changes in existing ontology structure or interpretation
  - Protracted discussion via SourceForge comments and email; resolve at face-to-face meeting
- *Example: pathogenesis & cell killing (four SourceForge entries, >30 emails, discussion at content meeting and after)*



# Documentation for Curators

- Existing: guidelines for
  - using SourceForge
  - editorial style & procedures
  - molecular function terms
- Forthcoming:
  - biological process term guidelines
  - cellular component term guidelines
  - decision-making history



# Documentation for Curators

GO Curator Guides

http://www.geneontology.org/GO.contents.curator.guides.html

## A Guide to Modifying the Gene Ontologies

These guides explain how to add to or alter the gene ontologies. They are intended for those members of the consortium whose job it is to make such modifications to the GO. However we make them available here for anyone who is interested to know how this work is carried out.

<a href="#">The Beginner's Guide to Modifying the Ontologies</a>	For curators who will be editing the ontologies. This guide is written to accommodate curators who may have only very basic computing experience, but also includes the information needed by more experienced users. It includes details of how to access the CVS repository.
<a href="#">Submitting suggestions to GO using SourceForge.net</a>	How to suggest possible changes to the ontologies.
<a href="#">Guide to Addressing a SourceForge Request</a>	How to make changes to the ontologies.

**Downloads**

- .. [Current Ontologies](#)
- .. [Current Annotations](#)
- .. [GO Database](#)

**Documentation**

- .. [FAQ](#)
- .. [Introduction](#)
- .. [Editorial Style Guide](#)
- .. [File Format Guide](#)
- .. [Function Ontology](#)
- .. [Synonym Guide](#)
- .. [Annotation Guide](#)
- .. [Evidence Codes](#)
- .. [SourceForge links](#)
- .. [Standard operating procedures](#)
- .. [Modifying GO](#)
  - ... [For Beginners](#)
  - ... [Suggest changes](#)
  - ... [Make changes](#)

[GO Tools](#)

[Mappings to GO](#)

**Archives**

[About GO](#)  
[Terms of Use](#)  
[Contact GO](#)  
[Report Errors](#)

**Search**  
Terms/Annotations