Ten years ago…

"The growth of scientific data, and of scientific databases, in the biomedical field—a growth not only of size but also of complexity—has been remarked upon so often as to become a cliché. The urge for 'integration' has been a mantra of both the bioinformatics community and of the funding agencies for decades."

"The growth of scientific data, and of scientific databases, in the biomedical field—a growth not only of size but also of complexity—has been remarked upon so often as to become a cliché. The urge for 'integration' has been a mantra of both the bioinformatics community and of the funding agencies for decades."

Integrating Biology

Towards a systems view

- 10 years ago:
  - What does this gene product do?

- Today:
  - What does it do in context?

We must do this efficiently

And make the most of the expertise
of our GO curators.

Increasing our efficiency

- Clear vision statement
- Responsibilities
- Communications
  - Talk to all affected parties, revisit these groups as each new milestone is reached
  - Use wiki and mailing lists effectively
- Establish milestones and dependencies
- Automate whenever possible

Ontology development
David Hill and Jane Lomax

- Role: Ensure that the Gene Ontology represents biology in a way that facilitates data integration.

Ontology development
Current responsibilities

- Written policies for developing ontology content
- New relationship specifications
- Daily updates to central ontology files
- Running logic checks on a regular basis and correcting any errors that are found.
- Making GO work with other ontologies
- Mappings

Ontology development Interactions

- Annotation: Incorporate feedback for clearer definitions, additional synonyms, and intuitive, self-evident naming conventions. Explain the relational structure of GO and its implications for annotation.

- RefG: Prioritize content development based on target annotation sets.

- Software: Assist in development of ontology reasoning tools, OBO-edit, automating ontology QC, the AmiGO team on effective ways to present multiple relationship types in the displays, and enhancement of other tools.

- External: Coordinate efforts in ontology development between GO and other relevant biological ontologies. Education emerging annotation groups on the use of GO.

Annotation Production: Rama Balakrishnan and Emily Dimmer

- Role: To ensure annotation quality, distribution, and effective use by the community.

Annotation Production
Current responsibilities

- Annotation training and education
  - Written guidelines for consistency in literature-curation.
  - Annotation training and consistency exercises.
  - Posting teaching/presentation materials on GOC web.
- Annotation QC
  - Enforce adherence to established GO standards and production policies (e.g. evidence codes, term usage…)
  - Run automatic annotation consistency checks.
  - Work with the software group to develop additional automated annotation checks wherever possible.
  - Incorporate feedback regarding annotations inconsistencies with respect to experimental annotations into the annotation guidelines and quality control checks

Annotation Production
Current responsibilities

- Annotation dissemination
    - Represent 'user perspective' in efforts to improve accessibility for all users.
    - Provide oversight of the go-help effort including tracking of user requests and development of FAQ for common questions,

Annotation Production Interactions

- Ontology: Provide feedback for GO content development.

- RefG: Document and develop quality assurance metrics to enforce the annotation consistency requirements.

- Infrastructure: Provide feedback on new features and bugs AmiGO, web site, downloads and other community user requirements

- External: Liaisons with MOD recommended experts, go-help desk, and web presence.

Reference Genome
Pascale Gaudet and Kara Dolinski

- Role: Complete annotation of 12 reference genomes:
    - Annotating with all available experimental data (breadth)
    - Annotating to the maximal feasible precision (depth)

Reference Genome
Current responsibilities

- Select annotation target sets for curation.
- Organize annotation topic meetings.
- Monitor progress towards comprehensive coverage
- Annotate by inference
- Deliver valid GAF files on a regular basis to MODs.
- Provide written annotation consistency recommendations to annotation managers
- Ensure that annotations are integrated by MODs
- Ensure gp2protein files are kept up-to-date

Reference Genome Interactions

- Ontology:
  - Provide recommendations for content development;
  - provide feedback on the ontology where the ontology structure (terms, definitions or relationships) make annotations ambiguous.
- Annotation:
  - Provide recommendations for improving annotation consistency.
  - Advise on community user requirements specific to reference genome data.
- Infrastructure:
  - Provide feedback on PAINT, database tracking.
- External:
  - Liaise with external experts, Reactome, and UniProtKB.

	January	February	March	March 30 April 1	April	May	June
PAINT (Ed, Suzi, Paul)		· Feb 1: beta17 release · Feb 9th: conference call: present beta17	· March 1: test release of V1 · March 15: public release of V1	GOC meeting SAB meeting	· April 12-13: Annotation meeting with PAINT, lung and tree people · Tree annotators give suggestions to GO camp session chairs for discussion topics	(software improvements)	
Lung annotation project (Li, Judy)	Last annotation targets selected						
Tree annotation (Kara, Mike, Pascale)		· Feb 1: test PAINT · Feb 15th: start releasing GAFs to MODs	Start annotation of lung targets				
MOD curators	Annotate lung targets. Priorities: (1) December targets; (2) January targets; (3) November targets				MOD curators review GAFs (that will be an ongoing process)		
GO camp (Pascale, Emily, Serenella)	Establish agenda		Session chairs start preparing materials for camp			May 15: each session chair gives 3 papers to camp participant to annotate	June 16-18: GO camp - Geneva
Curation report (Kara, Stan, Seth, Chris)		Feb 11: PANTHER families loaded in GO database	· March 1: Past family data loaded · March 25th: Status interface pages available for GOC meeting				

Technical Infrastructure
Chris Mungall (Suzi & Mike)

- Role: Provide the GOC and the user community with technical, software, bioinformatics, computer-science, and ontological reasoning support to create, clarify and extend the common model that underpins all of GO.

Technical Infrastructure
Current responsibilities

- Curation tools, both for ontology editing and ontology-based annotation.

- GO website infrastructure, AmiGO, query interfaces, timely database downloads, automated reports

- Software and analysis tools and application programmer libraries intended for bioinformatics and computational users

- Specification and documentation of core model used by by GO and it's various instantiations, including the file format specification for OBO and GAF formats, database schema, software object models, and ontology languages

- Creation of mirrors (EBI, Princeton) of the GO database

Technical Infrastructure Interactions

- Ontology:
  - Work with the ontology team on reasoning tools, OBO-edit development and enhancement of other tools.

- Annotation:
  - Provide support for AmiGO, the GODB, and other community user requirements

- ·RefG:
  - Provide support for PAINT and database tracking.

- External:
  - DB production, downloads, support requests, and working with external tool developers.

All of us

- <u>Publish!</u>
- Contribute to monthly reports
- Ask: how does this facilitate integration of biological data?