

# Core GO annotation activities

# Essential GO annotation activities for all groups

1. Annotations must be supplied in the primary **GAF 2.0 format** using either:

UniProtKB identifiers from the Reference Proteome Files

**OR** supply a complete gp2protein file to map from MOD specific identifiers

2. Provide a **complete, regularly updated gp2protein file when MOD ids used.**

Must contain :

- **all** gene products in the respective organism (or community) *including those with no GO annotation.*
- Protein coding genes must map to UniProtKB identifiers (Swiss-Prot in preference). If unavailable , mappings must be to NCBI identifiers (NP\_ and XP\_).
- If GO annotations are supplied for ncRNAs, then **all** ncRNA-encoding genes must be provided in a gp2rna file, mapping to NCBI ids. (RNACentral ids when available)
- Do not include heritable phenotypes/classical mutants

3. Support transition to using **ECO identifiers** in future annotation formats.

4. Where groups are unable to regularly release annotations sets, responsibility for updates will revert to the GOC.

## Responsibilities for the active GO Consortium groups

1. Be responsive to requests from other curators/external users for annotation corrections.
2. Be responsive in corrections uncovered from the GOC QC checks
3. Curators should be represented on the GO Consortium fortnightly annotation calls and frequently attend GO Consortium meetings.

## Responsibilities for the authoritative annotation source for a species

1. Integrate manual and electronic annotations from external sources (e.g. GOC, RefGenome, UniProt, Reactome) on a *monthly* basis. This should include data from the annotation\_extension field when provided.

# Activities expected of GO NIH funded curators

## Annotation activities:

1. Supply annotation files at least once a month.
2. Transition to using the more expressive annotation formats (e.g. GPAD/GPI)
3. Create highly informative annotations, including :
  - use specific protein forms (isoforms, post-translationally modified proteins), ncRNAs and protein complexes
  - biological context (spatial, temporal , functional and process contexts , target specificity)
  - annotation relations
4. Contribute to discussions on annotation expressivity enhancements e.g. LEGO-style annotation
5. Create annotations for *all* species described in a paper
6. Carry out phylogenetically-based annotations of PANTHER protein families , using PAINT

The above annotation activities will be supported by the Central GO Common Annotation Framework.

## Other activities:

7. Participate in development (e.g. testing) of the Central GO Common Annotation Framework, to ensure the provided software is available for use by all annotation groups in the GO Consortium
8. Mentor new GO annotation groups as allocated by the GOC.
9. Participate in the review of annotations contributed by external community experts