Current Development in the Evidence and Conclusion Ontology (ECO)

Rebecca Tauber*, James B. Munro, Suvarna Nadendla, Marcus C. Chibucos, & Michelle Giglio
Institute for Genome Sciences
University of Maryland School of Medicine
Baltimore, MD, USA
*rebecca.tauber@som.umaryland.edu

Abstract— The Evidence & Conclusion Ontology (ECO) has been developed to provide standardized descriptions for types of evidence within the biological domain. Best practices in biocuration require that when a biological assertion is made (e.g. linking a Gene Ontology (GO) term for a molecular function to a protein), the type of evidence supporting it is captured. In recent development efforts, we have been working with other ontology groups to ensure that ECO classes exist for the types of curation they support. These include the Ontology for Microbial Phenotypes and GO. In addition, we continue to support user-level class requests through our GitHub issue tracker. To facilitate the addition and maintenance of new classes, we utilize ROBOT (a command line tool for working with Open Biomedical Ontologies) as part of our standard workflow. ROBOT templates allow us to define classes in a spreadsheet and convert them to Web Ontology Language (OWL) axioms, which can then be merged into ECO. ROBOT is also part of our automated release process. Additionally, we are engaged in ongoing work to map ECO classes to Ontology for Biomedical Investigation classes using logical definitions. ECO is currently in use by dozens of groups engaged in biological curation and the number of ECO users continues to grow. The ontology, in OWL and Open Biomedical Ontology (OBO) formats, and associated resources can be accessed through our GitHub site (https://github.com/evidenceontology/evidenceontology) as well as the ECO web page (http://evidenceontology.org/).

Keywords—evidence; gene annotation; biocuration; ontology mapping; ontology development

I. INTRODUCTION

The Evidence and Conclusion Ontology systematically describes scientific evidence types in biological research. Biocurators, researchers, and data managers use evidence to support conclusions, such as an assertion that a protein has a particular function. ECO terms, as ontology classes, contain standard definitions and are networked with relationships. ECO is in use by numerous groups including large-scale resources such as UniProt-Gene Ontology Annotation (UniProt-GOA) which has >365 million evidence-linked GO annotations [1].

II. USER-DRIVEN DEVELOPMENT AND COLLABORATION

One of the core principles of the OBO Foundry is the "Commitment to Collaboration" [2]. Since our last publication in November 2016 [3] and the end of May 2018, over 1200 new evidence classes have been added to ECO. Most of our new class requests arrive directly from individual users via GitHub [4], but we work closely with various groups, highlighted below.

A. *Gene Ontology*

Since the origin of ECO, we have collaborated closely with the Gene Ontology (GO). The original set of ECO classes arose from the evidence codes used in GO annotations [1], and each GO evidence code has a corresponding ECO evidence class.

Recently, we added new GO-ECO mapped evidence classes to describe high throughput experiments, beginning with the Inferred from High Throughput Experiment (HTP) evidence code [5], mapped to 'high throughput evidence used in manual assertion' (ECO:0006056).

B. *Ontology for Microbial Phenotypes*

The Ontology for Microbial Phenotypes (OMP) has contributed numerous class requests to support their phenotype annotations. Currently, we have more than 68 classes traceable to OMP. Many of these terms fall under 'experimental phenotypic evidence' (ECO:0000059).

C. *Ontology for Biomedical Investigations*

We have been working with the Ontology for Biomedical Investigations (OBI) in a different scope than our other collaborators. Our goal is to harmonize ECO with OBI by adding logical definitions to ECO classes to describe how the evidence is generated during an investigation. While increasing the logic and consistency of the structure of ECO, this effort also benefits OBI by increasing their breadth of terminology.

We have added 41 new classes to OBI with 21 more pending additions. We have 188 ECO classes mapped to logical definitions created from OBI and GO classes. Our goal is to map all descendants of 'experimental evidence' (ECO:0000006).

III. ROBOT WORKFLOW

ROBOT is a command-line tool created to work with open biomedical ontologies [6], offering a series of commands to edit, review, and release ontologies. It is written in Java and the library is also available for programmatic use (http://mvnrepository.com/artifact/org.obolibrary.robot).

This material (the ontology & related resources) is based upon work supported by the National Science Foundation (NSF) Division of Biological Infrastructure (DBI) under Award Number 1458400 to MCC.

A. *Template-Generated Modules*

The *template* command takes a formatted spreadsheet (the template) and converts each row into one or more ontology axioms. This command has decreased the time it takes to create new classes, especially when creating multiple classes at once. Fig.1 shows a comparison of the standard workflow for creating classes in Protégé compared to creating classes with ROBOT.

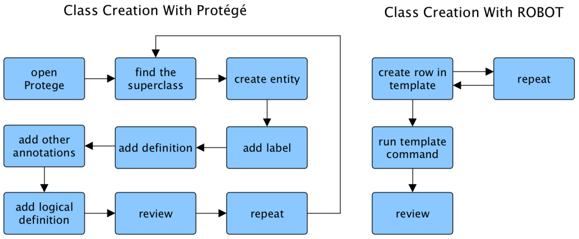

Figure 1. Class creation workflows

The *template* command lends itself easily to modular ontology development, which was first implemented by OBI [7]. The first module implemented by ECO is for the ECO-OBI harmonization project. Previously, the logical axioms were hard-coded into the *eco-edit.owl* file, causing inconsistencies due to human error. We are also adding modules for new class additions and "evidence used in assertion" classes.

B. *Import Management*

With our commitment to collaboration and ongoing ECO-OBI alignment project, we now import external ontology classes from OBI and GO into ECO. Both of these, especially GO, contain many unnecessary classes for our purposes, and importing the whole ontologies would be counterproductive. Therefore, we only wish to import a relevant subset of classes for our mappings using the ROBOT *extract* command. ROBOT offers the MIREOT [8] method of extraction so that our import modules have the necessary information to use the external classes.

By including *extract* as part of our release cycle, we regularly and automatically update the imported classes. Often, discrepancies are found between ontologies due to out-of-date imports, so this method of import management may be of use in other biomedical ontologies.

C. *Automating the Release Cycle*

In March 2018, we completed our first release using only ROBOT. Previously, a combination of tools (ROBOT, Protégé, OWLTools [9]) were used for different steps of the release. The release process also involved some manual editing to add version IRIs and timestamps. Now, the release process is entirely automated by various ROBOT commands in the Makefile.

Many biomedical ontologies already implement some form of a release workflow with a Makefile. Often, these workflows involve some degree of manual editing and review. ROBOT provides the framework to eliminate most of the need for this, and because it is a command line tool, it is straightforward to add to existing Makefiles.

IV. ONTOLOGY REVIEW

A. *Class Annotations*

Annotation axioms are those that describe a class to our users, such as the label, definition, and so on. These are important for our users to understand what the class is intended for and how to appropriately use it.

As of our last publication (November 2016), all classes had English labels, but almost 300 were missing definitions. In November of 2017, definitions were added for 99% of all classes. Going forward, we require a definition for all new term requests. We have also begun to annotate classes with their source using the 'ontology term requester' property. Having this level of provenance allows us to return to the requester if any changes need to be made, or if other issues arise.

B. *Logical Consistency*

ECO has been in development for well over 10 years [4], but there has not always been a standardized pattern for categorizing types of evidence. We have been reviewing classes, node by node, to ensure consistent categorization based on type of evidence rather than, for example, types of assays.

The goal of this review is to make it easier for users to find the correct class, and to not be confused by similar and ambiguous classes. It also sets a standardized pattern for adding new evidence classes in the future.

V. FUTURE WORK

We will continue to collaborate closely with GO, OMP, OBI, and other groups, as well as monitor GitHub requests [3]. Finally, we will continue our internal projects to increase logic between ECO classes by adding logical definitions, complete the OBI mappings, and review the categorization of evidence.

ACKNOWLEDGMENT

The ECO working group thanks our collaborators and our entire user-base for continuing to help us grow.

REFERENCES

[1] M. C. Chibucos, et al., "Standardized description of scientific evidence using the Evidence Ontology (ECO)," Database, vol. 2014, January 2014.

[2] The OBO Foundry, "Principle: Commitment To Collaboration," http://obofoundry.org/principles/fp-010-collaboration.html.

[3] M. C. Chibucos, D. A. Siegele, J. C. Hu, M. Giglio, "The Evidence and Conclusion Ontology (ECO): Supporting GO annotations," in The Gene Ontology Handbook, C. Dessimoz, N. Skunca, Eds. pp.245-259.

[4] "Issues – evidenceontology," https://github.com/evidenceontology/evidenceontology/issues.

[5] Gene Ontology Consortium, "Inferred from high throughput experiment (HTP)," http://wiki.geneontology.org/index.php/Inferred_from_High_Throughput_Experiment_(HTP).

[6] "ROBOT is an OBO Tool," http://robot.obolibrary.org/.

[7] "The Ontology for Biomedical Investigations," https://github.com/obi-ontology/obi.

[8] M. Courtot, et al., "MIREOT: the Minimum Information to Reference an External Ontology Term," Nature Preceedings, August 2009.

[9] "OWLTools," https://github.com/owlcollab/owltools