

bio.tools studentship

Annotating software tools in domains of the Life Sciences

Author(s): Veit Schwämmle, Jon Ison. **Contributors:** Michael Crusoe

<http://tinyurl.com/biotoolsstudent4>

The ELIXIR Tools and Data Services Registry [1] (<https://bio.tools>) aims to provide comprehensive coverage of biological research software tools and online services built upon the primary biological databases. To fulfill the registry mission - to boost tool and data utilisation - it is essential to adequately catalogue these software. For biological research software to be registered in bio.tools, certain semantic annotations are mandatory, as specified in the biotoolsSchema [2] and defined in terms from the EDAM ontology [3]. EDAM provides four categories of terms for precise annotation and characterisation of tools within the Life Sciences: *topic* (a broad category within the life sciences), *operation* (specifically what a tool does), types of input and output *data*, and data *format*. bio.tools aims to offer not only a comprehensive catalog of current biological research software tools but also provide easily accessible information for the identification of task-specific software, and for facilitating usage including integration into computational pipelines. Such an ambitious goal requires extensive and continuous efforts for the correct and sufficiently comprehensive annotation of thousands of computational tools, including improvement and extension of the EDAM ontology.

Task: We offer student projects within different thematic areas of the Life Sciences to achieve more representation in bio.tools with richer, more consistent annotations. The task involves combining computer-aided annotation of software tools, manual editing of bio.tools entries including correction of incorrectly assigned descriptors, alongside development of new terms as necessary for inclusion into the EDAM ontology.

We offer studentships with a focus on any specific domain of the Life Sciences; priority will be given to current research themes and areas that are not currently well represented in bio.tools. Depending on the state of coverage and annotations in bio.tools in the assigned theme, the tasks include:

- mine the literature for available software and gather detailed information about these tools including functionality, file input and output and specific data formats -- proposing new EDAM terms or revisions to existing EDAM terms as necessary for adequate description of software
- register individual tools or (programmatically) transform and export information about existing tool collections *en masse*, assigning EDAM terms and improving the annotation to a high standard
- review and improve existing bio.tools registrations within the theme

Students are encouraged to apply their expertise and gathered knowledge in ways that directly and positively impact the growth and quality of bio.tools. Students will simplify and accelerate the curation process, including the development of appropriate computational utilities. The projects therefore require independent work and innovation.

Requirements: Students must be familiar with concepts in a scientific area, highly self-motivated with fine attention to detail, proficient in English, and motivated to invest time and efforts into high-quality characterization of biological research tools.

Expected outcome: Students will gain in-depth knowledge about bioinformatics tools and their categorization, and will learn to assess the scope and utility of Life Science software. Further involvement in these highly relevant projects can be continued and most likely will result in a publication in an international peer-reviewed journal.

Supervision: All Supervision will be in English and by expert editors of the scientific theme.

Contact: Veit Schwämmle (veits@bmb.sdu.dk) cc Jon Ison (jison@cbs.dtu.dk) and Henriette Husum Bak-Jensen (hhu@bio.ku.dk).

[1] Ison, J. *et al.*, 2015, Tools and data services registry: a community effort to document bioinformatics resources. *Nucleic Acids Research*, 44(D1): 38-47. doi: 10.1093/nar/gkv1116

[2] <https://github.com/bio-tools/biotoolsschema>

[3] Ison, J. *et al.*, 2013, EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, 29(10):1325-1332. doi: 10.1093/bioinformatics/btt113