

# bio.tools studentship

## Annotating software tools in a scientific context

**Author(s):** Jon Ison. **Contributors:** n/a

<http://tinyurl.com/biotoolsstudent3>

The ELIXIR Tools and Data Services Registry [1] (<https://bio.tools>) aims to provide rich descriptions of prevalent biological research software tools and online services built upon the primary data resources. Central to the aim of providing a world-leading registry of software information, the registry must place software in the scientific context, so that a scientist who is searching for new tools can find the best tools for the job, explore the latest offerings, and understand the context in which tools are being developed and applied. This aim is partly fulfilled by annotating tools in terms from the EDAM ontology [2] which provides a precise nomenclature for annotation and characterisation of tools within the Life Sciences: *topic* (a broad category within the life sciences), *operation* (specifically what a tool does), types input and output *data*, and data *format*. The second way the aim is fulfilled is to cross-link bio.tools with the scientific literature. For this purpose, biotoolsSchema [3], the formal data model upon which bio.tools is based, specifies a link for publications that may be annotated according to publication type (see below). Secondly, a prototype new interface [4] includes figures extracted from tool publications and converted into interactive diagrams.

**Task:** we offer a student project with a focus on tasks around integration of bio.tools with the scientific literature:

- systematically annotate existing bio.tools entries with publications, using existing publication lists where available<sup>1</sup>, marking these as per biotoolsSchema as one of “Primary”, “Benchmark”, “Review”, “Other”<sup>2</sup>
- create and optimise appropriate literature searches to identify list of new software tools published in the literature, cross-referencing the lists to bio.tools to identify tools which are not yet registered
- work with the bio.tools developers to expose within bio.tools metrics and and other information from literature analysis including number of citations, Altmetric rating, bookmarklets, publication summary *etc.* building upon the ongoing discussions<sup>3,4,5</sup>
- *at a lower priority ...* create new interactive tool diagrams from publications available under open license

Students are expected to work in a highly focussed way that directly and positively impact the growth and quality of bio.tools. The projects require independent work and innovation is welcome upon ways to deepen and improve the integration of bio.tools with the literature.

**Requirements:** Students must be familiar with scientific software literature, highly self-motivated with fine attention to detail, proficient in English, and motivated to invest time and efforts into high-quality curation of bio.tools.

**Expected outcome:** Students will gain in-depth knowledge about bioinformatics tools and will gain valuable first hand experience of working in a busy production environment. Continued involvement in bio.tools is welcomed and can results in co-authorship of a publication in an international peer-reviewed journal.

**Supervision:** All Supervision will be in English (supervisors tbd).

**Contact:** Jon Ison ([jison@cbs.dtu.dk](mailto:jison@cbs.dtu.dk)) cc Henriette Husum Bak-Jensen ([hhu@bio.ku.dk](mailto:hhu@bio.ku.dk)) and Hans Ienasescu ([hans.ienasescu@gmail.com](mailto:hans.ienasescu@gmail.com))

[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

[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

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

[4] <https://bio.tools/workflows>

---

<sup>1</sup> <https://biotools.sifterapp.com/issues/356>

<sup>2</sup> this classification is under review and may be refined, see <https://github.com/bio-tools/biotoolsSchema/issues/73>

<sup>3</sup> <https://github.com/bio-tools/biotoolsregistry/issues/106>

<sup>4</sup> <https://github.com/bio-tools/biotoolsregistry/issues/60>

<sup>5</sup> <https://github.com/bio-tools/biotoolsregistry/issues/35>