bio.tools & EDAM
drop-in hackathon & discussions

- EDAM and bio.tools developers are here all week
- different theme each day, but flexible (just drop by)
- suitable for anyone interested to learn about, use or develop EDAM and bio.tools
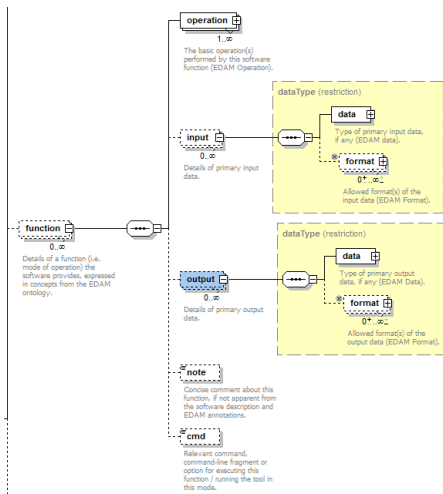
Jon Ison

Hans-Ioan Ienasescu

Matúš Kalaš

Hervé Ménager

Veit Schwämmle

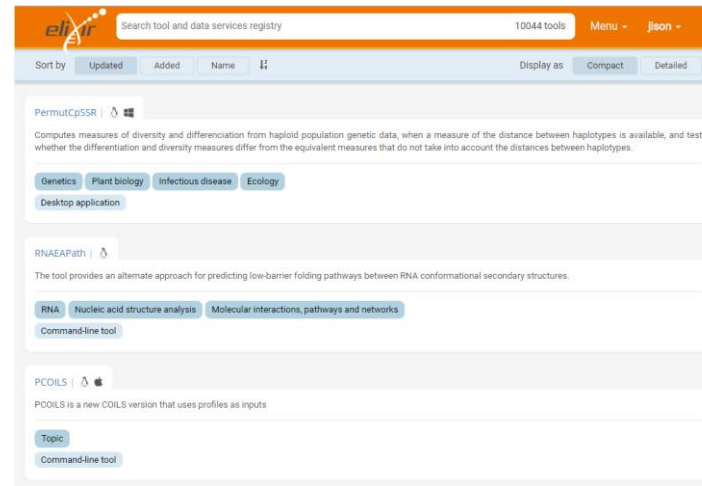# biotoolsSchema

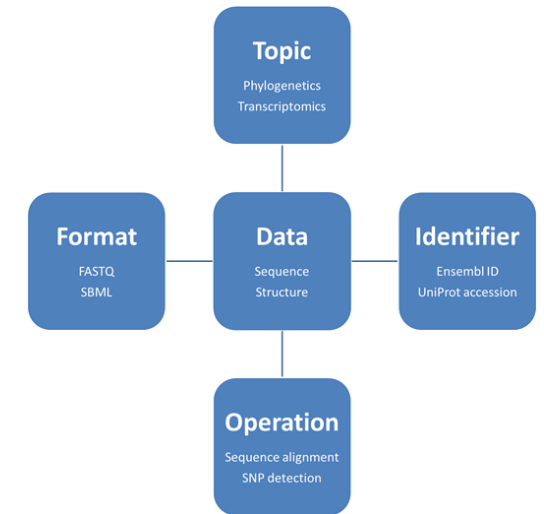github.com/bio-tools/biotoolsschema

formalised schema (XSD)
50 software attributes
16 controlled vocabularies
rigorous **syntax**

bio.tools

https://bio.tools

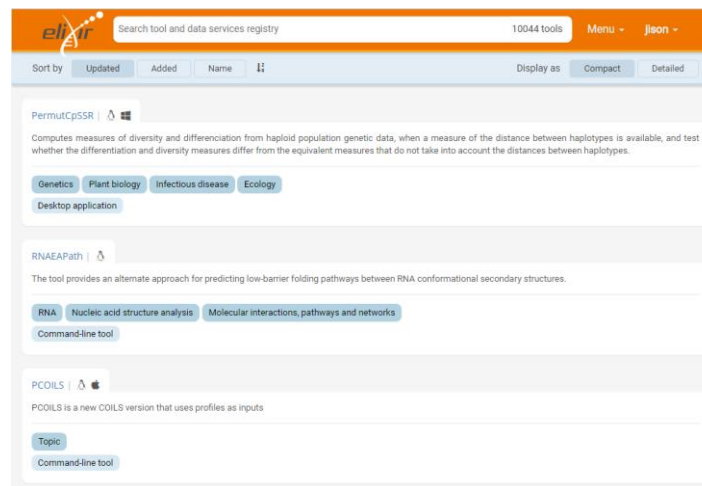*registry of life-science software*

EDAM ontology

github.com/edamontology/edamontology

controlled vocabulary
topics, operations, types of data, data
identifiers, data formats
3,354 concepts
rigorous **semantics**

Tool Information Standard

bio-tools.github.io/Tool-Information-Standard/

"min/max" information for a tool
what attributes should be specified
at various levels of description
completeness

bio.tools

https://bio.tools

registry of life-science software

Curation Guidelines

biotools.readthedocs.io/en/latest/
curators_guide.html

best-practice guidelines
human-readable
attribute & tool-specific guidelines

Tool Cards at persistent URLs

Unique tool identifiers

12,205 entries

227,752 annotations

940 contributors

10K user visits / month

open data (CC-BY-4.0)

open software (GPL 3.0)
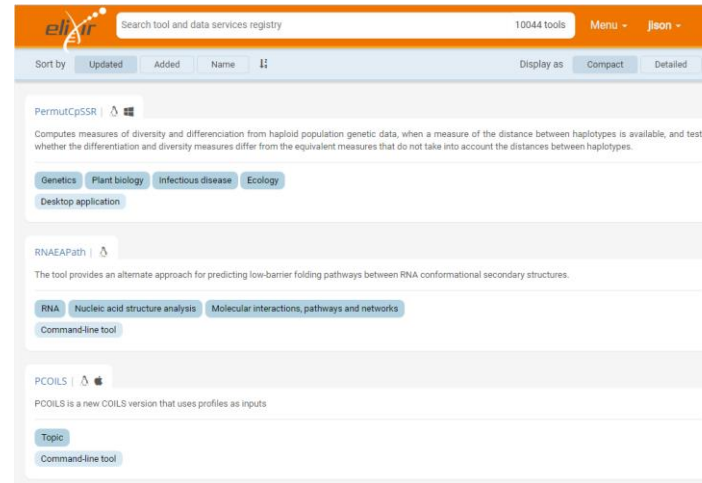
bio.tools

https://bio.tools

registry of life-science software

latest (much better) version
dev.bio.tools

bio.tools API
bio.tools/api/tool

extensive docs
biotools.readthedocs.io/en/latest/

get involved
github.com/bio-tools/biotoolsregistry/

Day 2 (Nov 13): bio.tools testing

Expected audience: *anyone with an interest in improving bio.tools*

Expected outcome: *verify the next release, improve the search performance*

The purpose is to test, evaluate and optimise the development deployment of bio.tools (https://dev.bio.tools/), changes in which are scheduled to be moved into production (https://bio.tools/) during Dec 3-7. The bio.tools core-dev will be on hand to discuss things in person.

Task 1: Release testing

Currently 28 issues labelled "done - staged for release" are implemented in https://dev.bio.tools. Before these can be moved into production, we need independent verification that these features and fixes are satisfactorily implemented.

The task is:

- pick any "done - staged for release" issue which lacks the "fix verified" label
- read the thread and test things are working as advertised
- add a comment to the thread; either reporting things are OK, or describing an outstanding problem: bio.tools core-dev will monitor the tracker, fix issues that crop up, and attach the `fix verified` label to confirmed fixes
- repeat, until all `done - staged for release` issues are verified

Alternatively:

- experiment with https://dev.bio.tools - critique the interfaces, API and content - and report any bugs or suggestions via GitHub

Task 2: bio.tools API testing & optimisation

The latest development deployment of the bio.tools API (https://dev.bio.tools/api/tool) is, we hope, a big improvement on the current version. It supports a comprehensive set of parameters that enable precise query over tool function and other metadata. But before we can move these changes into production, the API needs to be thoroughly tested. We also want to optimise the search behaviour, in light of results of real user experiments, to ensure it works as anticipated.

The task is:

- systematically test the API, particulaly the behaviour of the search parameters as documented in the API Reference and API Usage Guide.
- provide feedback on the API search behaviour / possible improvement via GitHub. You can suggest fixes or improvements to the API docs here.
- *elasticsearch experts only* - please speak to bio.tools core-dev (there are issues we need help with!)

We hope (developments pending) to have an easy way to tweak the elasticsearch parameters during the workshop, allowing for immediate iterative improvements.

Day 3 (Nov 14): bio.tools outreach

Expected audience: *anyone with an interest in developing bio.tools*

Expected outcome: *kick-start the community development process*

The purpose is introduce our current development priorities and to introduce and improve the proposed community development process for bio.tools. The bio.tools core-dev will be on hand to discuss things in person.

Task 1: Development priorities

We label issues to reflect their status and priority:

- "critical priority" : our top priorities, including most of the reported bugs
- "high priority" : things which bio.tools core-dev consider high priorities; we get to these once "critical priority" i addressed
- "in progress" : things we're working on currently
- "Dec 18 release" : things we're aiming to put into the next production deployment
- "wontfixsoon" : things which, for one reason or another (usually lack of developer capacity), we don't anticipate doing soon (that's doesn't imply they're unimportant or bad ideas!)

We want to be sure our priorities reflect those of the community at large, and engage developers who are willing to help out. The task is:

- review our priorities (issues in any of the categories above) - providing feedback in the appropriate GitHub thread
- feel free to request new features, but please first search our issues as it might already be listed
- *developers only* - if you're interested to help out - especially on "critical priority" issues (or anything else!), then please disucss this with the bio.tools core-dev

Task 2: Open development process

Now that bio.tools is open source, there is an opportunity for hackers everywhere to contribute to the project. But first we must define how the community development process will work in practice. We have emerging contributor guidelines but we want to revise these in light of feedback from potential contributors.

The task is to review the emerging contributor guidelines, provide feedback on these via GitHub, or provide feedback in person to bio.tools core-dev.

Day 4 (Nov 15): EDAM development

Expected audience: *anyone with an interest in improving EDAM, people knowledgeable of bioinformatics data formats*

Expected outcome: *improved EDAM Formats subontology, scoping the desired state of EDAM 2.0, developing EDAM applications*

Task 1 Curation of bioinformatics data formats

The EDAM Format subontology has potential in systems such as Galaxy and for applications such as workflow composition. EDAM is close to providing a comprehensive catalogue of the prevalent bioinformatics data formats, but a significant amount of work remains. The task is to work on any aspects of the data format curation listed here including:

- addition of miscellaneous new data formats, or changes to existing ones (see issues)
- addition of formats ensuring coverage for Galaxy applications (issue)
- addition of formats to ensure coverage of FAIRSharing

We expect the tasks to be accomplished manually, programmatically, or by a combination of the two. Please see:

- guidelines for EDAM Format curation
- EDAM Format request template (see the docs)

Task 2 Verification of EDAM Formats subontology

We have guidelines for the development of the EDAM formats subontology:

- editor guidelines when modifying EDAM; adding or changing concepts, concept metadata, crosslinking, etc.
- developer guidelines about the technical process

To develop EDAM Format subontology into a rigorous catalogue, we must ensure the guidelines are followed. The task is:

- review the editor guidelines and developer guidelines, and provide feedback on these via GitHub or discuss this in person with EDAM core-dev
- propose clean-ups of the connection between EDAM Format and Data subontologies (see issue) : please make suggestions via GitHub - see also issue
- (*developers only*) develop a utility that checks compliance of EDAM to the guidelines above, and generate a human-readable report that can be acted on. In case you want to work with EDAM in JSON / JSON-LD format, see edam2json

Task 3 Towards EDAM 2.0 (discussion & planning)

It's over 5 years since an article describing EDAM was published in Bioinformatics. Since then, there have been 18 new releases (currently EDAM 1.21), with many additions and improvements, and greatly improved documentation:

- rules for concepts,
- rules for hierarchy depth
- guidelines for editors
- guidelines for developers

Within 3 - 6 months, we hope to release EDAM 2.0 implementing a set of features representing a step forward in value and quality over the 1.* releases. The task (working as a group, or alone) is:

- think; what are the desirable properties of EDAM 2.0? Is it simply to adhere to the rules and guidelines above, or something more?
- enumerate desirable properties in this issue; we'll try to prioritise these during the hackathon
- create sub-issues as needed, for finer-grained information

Task 4: EDAM applications (discussion & hacking)

EDAM is used (or being considered) in a variety of contexts. There is an opportunity for developers on projects that are using (or considering) EDAM to discuss their requirements and work with the EDAM developers. Or you might have an idea that we haven't heard of already; let's discuss.

Day 5 (Nov 16): Planning & coordination

The final day will be reserved to finishing off, and discussing and planning next steps around collaborations of EDAM and bio.tools with other projects.