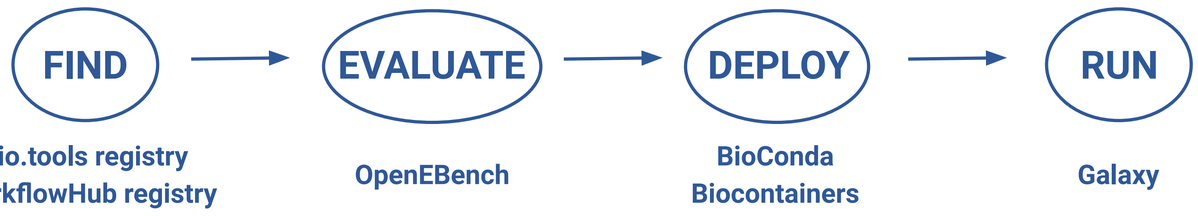
The ELIXIR Tools Ecosystem

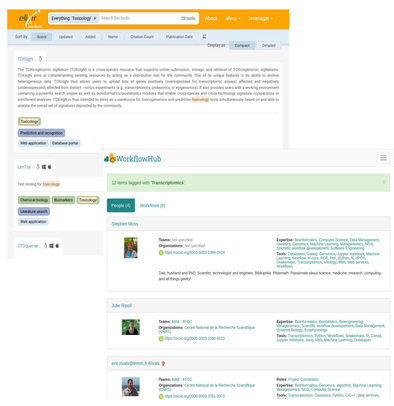
Salvador Capella-Guttierez[1], Björn Grüning[2], Johan Gustafsson[3], Matúš Kalaš[4], Hervé Ménager[5], Tools Ecosystem Development Team[6]

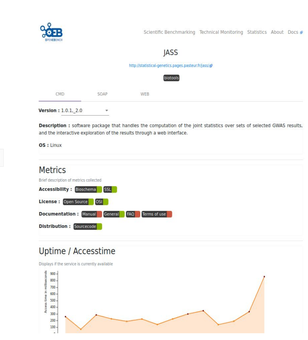
ELIXIR All Hands , 5-8 June 2023, Dublin, Ireland

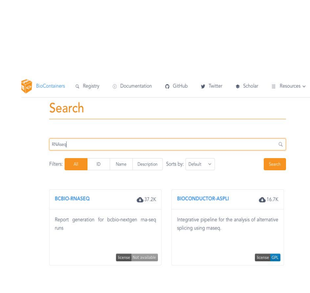
Research software is a critical component of computational research. Thus, being able to discover, understand and adequately utilize software is essential. These activities all rely on software metadata metadata. However, these metadata are sparse, expensive to maintain, and frequently inconsistent across different resources.



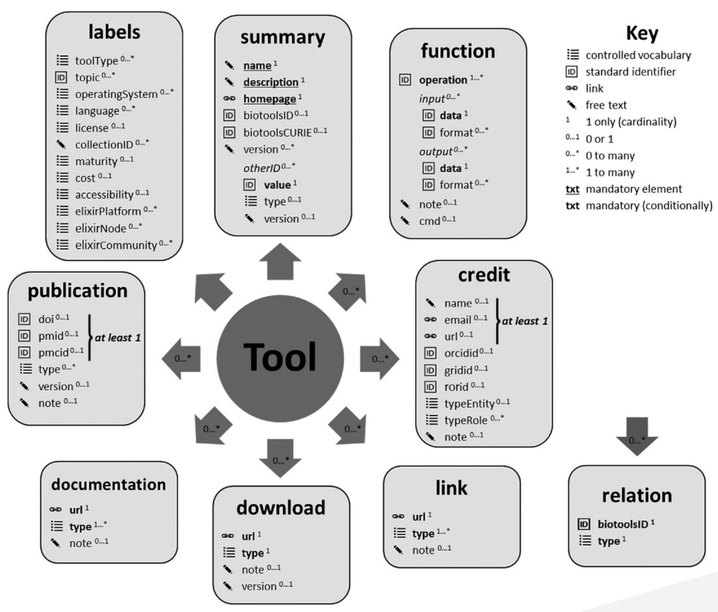
The ELIXIR Tools Ecosystem acts as a proxy to maintain and preserve high-quality metadata for describing research software, aggregating multiple heterogeneous metadata resources. These resources often have different objectives, user communities and technical implementations.

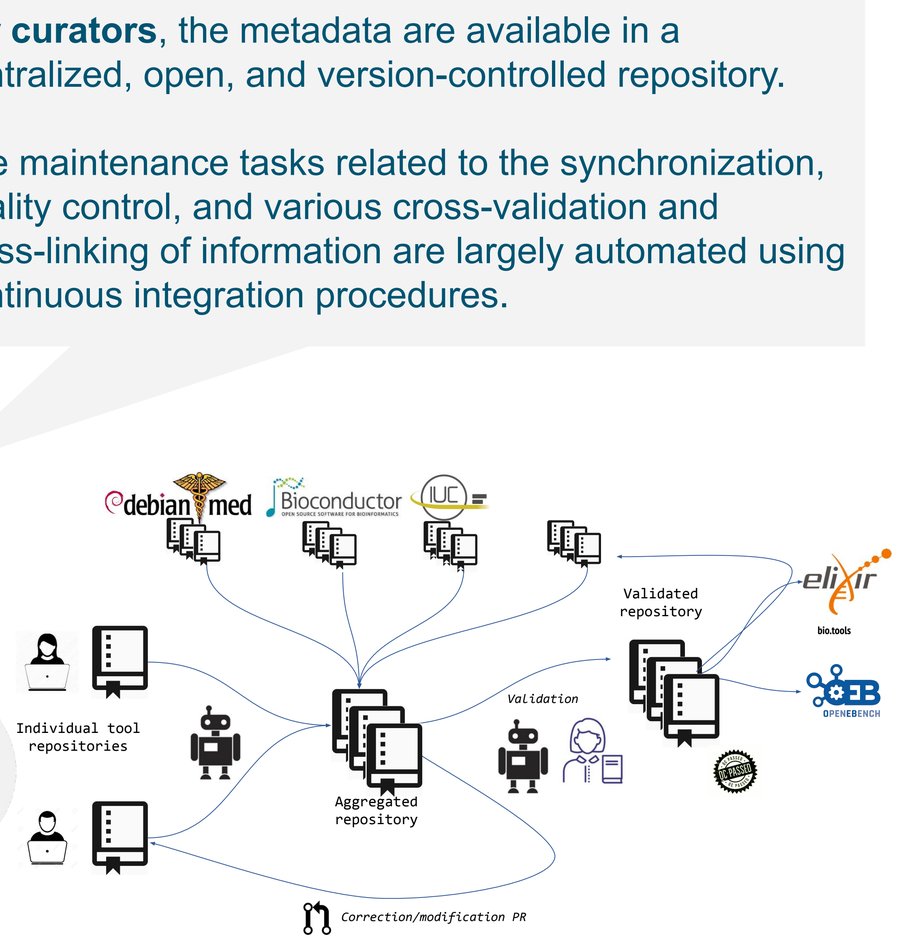






The goals of this project are to consolidate these metadata, and facilitate their integration, curation, and re-usability by the contributors and the wider research community.





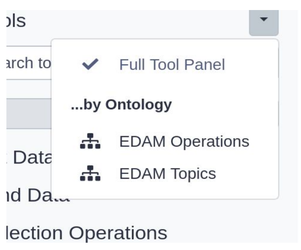
For **developers**, the ecosystem facilitates the description of their software and workflows. It centralizes all the available metadata related to a given software in their source format, but the use and mapping of standards such as EDAM, biotoolsSchema and BioSchemas will facilitate the propagation of these metadata, and improve their overall FAIRness.

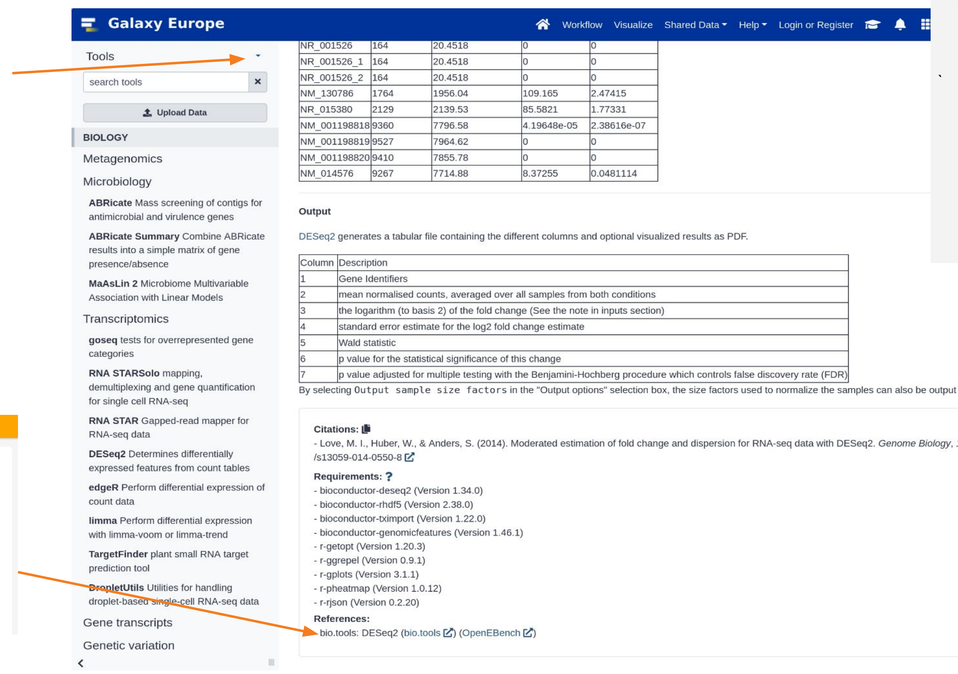
For **curators**, the metadata are available in a centralized, open, and version-controlled repository.

The maintenance tasks related to the synchronization, quality control, and various cross-validation and cross-linking of information are largely automated using continuous integration procedures.

For **users**, this resource is directly queried by multiple systems, such as:

- The Australian BioCommons ToolFinder, which lists bioinformatics tools and software that are installed across several of the BioCommons partners.
- Galaxy, which uses it e.g. to sort tools according to their EDAM annotations, or to provide links to the corresponding bio.tools entries.





NR_001526	164	20.4518	0	0
NR_001526_1	164	20.4518	0	0
NR_001526_2	164	20.4518	0	0
NM_130786	1764	1956.04	109.165	2.47415
NR_015380	2129	2139.53	85.5821	1.77331
NM_001198818	9360	7796.58	4.19648e-05	2.38616e-07
NM_001198819	9527	7964.62	0	0
NM_001198820	9410	7855.78	0	0
NM_014576	9267	7714.88	8.37255	0.0481114

Output

DESeq2 generates a tabular file containing the different columns and optional visualized results as PDF.

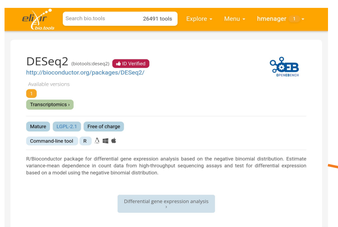
Column	Description
1	Gene Identifiers
2	mean normalised counts, averaged over all samples from both conditions
3	the logarithm (to basis 2) of the fold change (See the note in inputs section)
4	standard error estimate for the log2 fold change estimate
5	Wald statistic
6	p value for the statistical significance of this change
7	p value adjusted for multiple testing with the Benjamini-Hochberg procedure which controls false discovery rate (FDR)

By selecting `Output sample size factors` in the "Output options" selection box, the size factors used to normalize the samples can also be output as a tabular file.

Citations:
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Requirements:
- bioconductor-deseq2 (Version 1.34.0)
- bioconductor-rhdf5 (Version 2.38.0)
- bioconductor-tximport (Version 1.22.0)
- bioconductor-genomicfeatures (Version 1.46.1)
- r-getopt (Version 1.20.3)
- r-ggrepel (Version 0.9.1)
- r-gplots (Version 3.1.1)
- r-pheatmap (Version 1.0.12)
- r-rjson (Version 0.2.20)

References:
bio.tools: DESeq2 (bio.tools) (OpenEBench)



For **users**, this resource is directly queried by multiple systems, such as:

- The Australian BioCommons ToolFinder, which lists bioinformatics tools and software that are installed across several of the BioCommons partners.
- Galaxy, which uses it e.g. to sort tools according to their EDAM annotations, or to provide links to the corresponding bio.tools entries.

[1] Barcelona Supercomputing Center, Barcelona, Spain.
[2] University of Freiburg, Freiburg, Germany.
[3] Australian Biocommons, University of Melbourne, Australia.
[4] University of Bergen, Bergen, Norway.
[5] Institut Pasteur, Paris, France.
[6] ELIXIR Europe, Many Places

Project Website: https://elixir-europe.org/internal-projects/commissioned-services/tools-platform-ecosystem
Source Code: https://github.com/research-software-ecosystem/
License:
Code (https://github.com/research-software-ecosystem/utils) MIT
Content (https://github.com/research-software-ecosystem/content) CC-BY