

The RNA tools registry

A community effort to catalog RNA bioinformatics resources
and their relationships

Anne Wenzel

Center for non-coding RNA in Technology and Health
Faculty of Health and Medical Sciences
University of Copenhagen

32nd TBI Winterseminar
Bled, Slovenia
February 2017



Bioinformatics tool interoperability and integration

- ▶ Creation of a comprehensive tool registry
- ▶ Provision of tools in the form preferred by users across academia and industry
- ▶ Benchmarking, tool sustainability and renewal
- ▶ Tool interoperability

- ▶ ELIXIR Tools and Data Services Registry
- ▶ Collection of bioinformatics resources
- ▶ 5226 entries currently (goal: 10 000 by the end of 2017)
- ▶ Mandatory attributes:

Name

Version

Description Free text

Function Ontology term

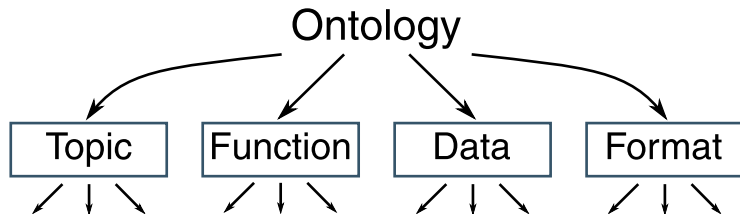
Topic Ontology term

Homepage URL

Contact Email, ...

Tool Type [Command-line tool, Database portal,
 Desktop application, Web service, ...]

*(Ison *et al.*, *Nucleic Acids Res*, 2016)



Topic is a broader domain or field of interest, of study, data, or technology, e.g., 'MicroRNA', 'RNA-seq'

Function describes what a tool does, but not (typically) how, e.g., 'RNA-RNA interaction prediction'

Data describes the type of (input or output) data, e.g., 'RNA sequence' or 'Multiple RNA sequence alignment'

Format of the *Data*, e.g., 'FASTA' or 'Stockholm'

* (Ison *et al.*, Bioinformatics, 2013)

Search for RNA interaction in EDAM...

Preferred Name	Protein interactions
Synonyms	Protein-nucleic acid interactions Protein-RNA interaction Protein-ligand interactions Protein-protein interactions Protein interaction networks Protein-DNA interaction
Definitions	Protein-protein, protein-DNA/RNA and protein-ligand interactions, including analysis of known interactions and prediction of putative interactions. This includes experimental (e.g. yeast two-hybrid) and computational analysis techniques.
ID	http://edamontology.org/topic_0128

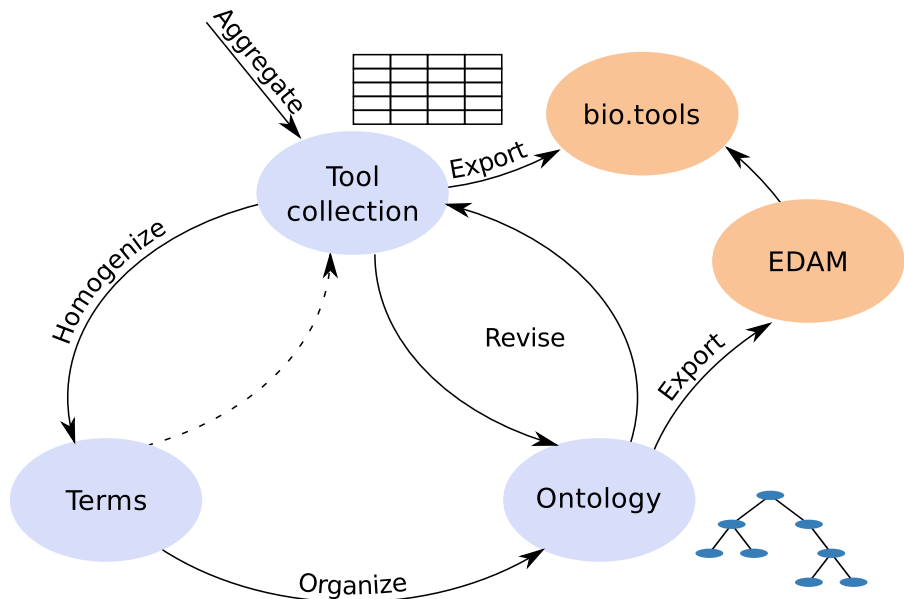
... only reveals Protein-RNA interaction

Curatathon I



Photo: T. Houwaart

Process outline



Curatathon II



Photo: N. Doncheva

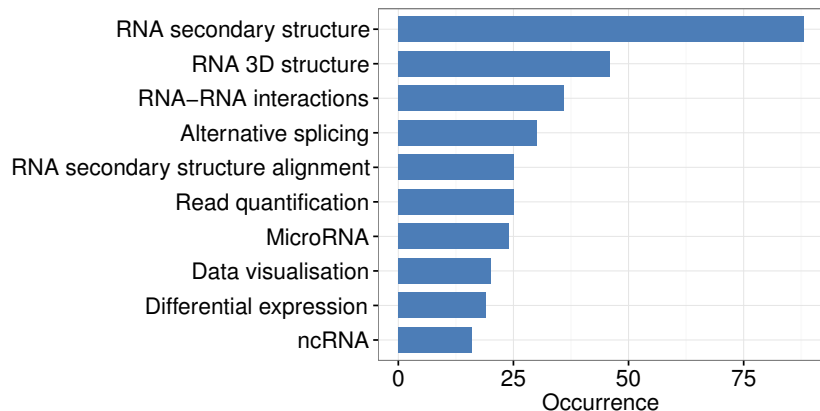
Since then ...

- ▶ Added 20 tools
- ▶ Automatic validation checks
- ▶ Adapt to new bio.tools schema
- ▶ Further sanitizing entries
- ▶ Carefully hand-curated 90 entries

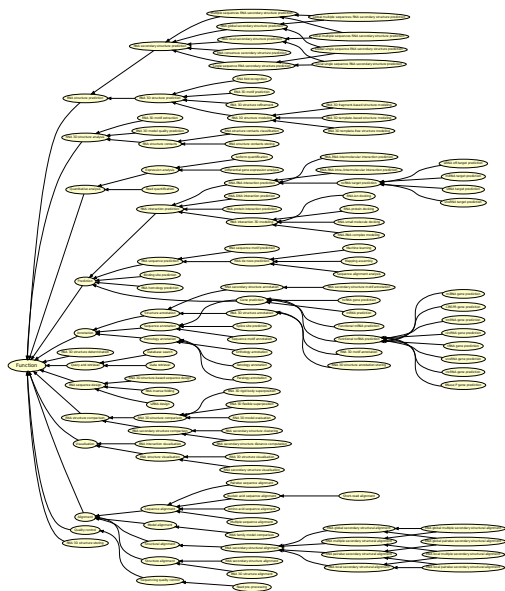
Results: Contents of the registry

389 tools currently in the registry

Ten most frequently used *Topic* terms:



Results: RNA tools' *Function* Ontology



DEMO

Conclusion

For the individual resources, we provide

- ▶ comprehensive description
- ▶ systematic classification
- ▶ relations to other resources

Conclusion

For the individual resources, we provide

- ▶ comprehensive description
- ▶ systematic classification
- ▶ relations to other resources

To be integrated in

- ▶ EDAM : ontology
- ▶ bio.tools : registry

Conclusion

For the individual resources, we provide

- ▶ comprehensive description
- ▶ systematic classification
- ▶ relations to other resources

To be integrated in

- ▶ EDAM : ontology
- ▶ bio.tools : registry

Allowing a researcher to

- ▶ identify the right functional class to solve the problem at hand
- ▶ find alternatives for certain tools

Conclusion

For the individual resources, we provide

- ▶ comprehensive description
- ▶ systematic classification
- ▶ relations to other resources

To be integrated in

- ▶ EDAM : ontology
- ▶ bio.tools : registry

Allowing a researcher to

- ▶ identify the right functional class to solve the problem at hand
- ▶ find alternatives for certain tools

Our approach is relevant for other community efforts

- ▶ Graphical ontology-driven entry point to the tools registry
- ▶ Open for contributions by 3rd parties
- ▶ Text-mine abstracts (using existing edamMap)
- ▶ Journal incentive to require accession before publication
- ▶ Interfacing with RNAcentral

Acknowledgements

The **RNA tools and software consortium** consists of researchers from:

University of Vienna, Austria

University of Freiburg, Germany

University of Jena, Germany

University of Leipzig, Germany

European Bioinformatics Institute, Hinxton, UK

International Institute of Molecular and Cell Biology, Warsaw, Poland

Paris-Sud University, Orsay, France

École polytechnique, Palaiseau, France

University of Copenhagen, Denmark

We acknowledge the Danish infrastructure roadmap grant to the Danish ELIXIR node and Innovation Fund Denmark (0603-00320B) to the grant for RTH.