# Better findability for BioC packages

through semantic annotation

## How we got here

Some people wrote a review The metaRbolomics Toolbox in Bioconductor and beyond of metabolomics packages in R, which started as an idea to show how BioC packages in biocView Metabolomics work together.

Turned out we found *FAR* more packages than expected, in BioC and beyond. Putting everything together was tough (classic) literature research.

## Findability of R packages

In some places in the article we mention findability:

- Section “1.2. The R Package Landscape” describes “CRAN Task Views”, “BiocViews”
- https://rdrr.io/ is a comprehensive index of R packages and documentation from CRAN, Bioconductor, GitHub and R-Forge, in “3. Conclusions” we mention that GitHub has a concept of topics: https://github.com/search?q=topic:metabolomics+topic:r
- A Fun exercise was to create Figure 2 which revealed even more metaRbolomics packages on the way, by connecting dependencies from DESCRIPTION

## What would be even cooler

Numerous times we sighed and wondered why it isn’t easier to find, navigate and classify R packages.

### Adoption of Bioschemas

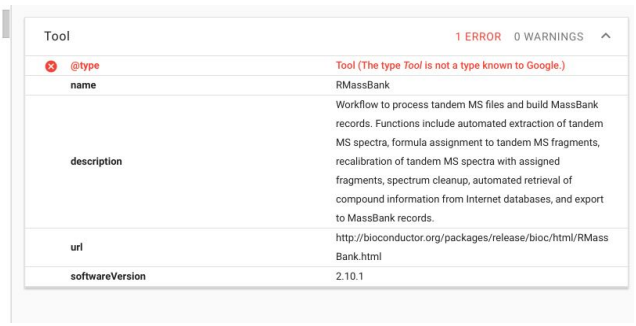
Bioschemas is about Semantic annotation invisible (to humans) inside HTML:

1. Bioschemas is a community project built on top of schema.org, aiming to improve interoperability in Life Sciences so resources can better communicate and work together by using a common markup on their websites.
2. Bioschemas reuses terms from well-known ontologies thus avoiding reinventing the wheel. `Tools`, a `SoftwareApplication` profile, recommends using terms from the EDAM Ontology (browse in bioportal.....org/.../EDAM or ebi.ac.uk/ols/.../edam)

See also the Bioschemas paper and tutorial.

## Bioschemas in BioC Websites powered by DESCRIPTION

Expose content from the DESCRIPTION file as Bioschemas annotations on Bioconductor by adding to the BioC Website templating in
github.com/.../bioconductor.org/.../_bioc_views_package_detail.html
https://github.com/Bioconductor/bioconductor.org/pull/25


## Bioschemas in Vignettes

Egon Willighagen looked into BioSchemas annotation for tutorials (`CreativeWork`) and tested that with the BridgeDbR package, and the results of that is written up in this blog post:
https://chem-bla-ics.blogspot.com/2019/04/bioschemas-creativework-annotation-in.html

Efforts to start annotation in vignettes allows the ELIXIR Training eSupport System TeSS (https://tess.oerc.ox.ac.uk) to pick up training material from bioconductor.org/.../vignettes/BridgeDbR/.../tutorial.html (source in BridgeDbR_vignette) through a sitemap.xml which is registered in TESS resulting in tess.elixir.org/materials?tools=BridgeDb

# The Elixir bio.tools registry

bio.tools/ strives to provide a comprehensive registry of software and databases from simple command-line tools [...] to complex, multi-functional analysis workflows. Resources are described in a rigorous semantics and syntax.

- Example for a (manually) well-done entry for a single tool: bio.tools/jmztab-m
- Query all R packages on Metabolomics: bio.tools/t?topic=Metabolomics&language=R
- There is a machine-readable API: bio.tools/api/t/?biotoolsID=”xcms”
- And there is support & tooling for mass-importing packages:
  R/CRAN/BioC content import documentation and policy
- Sidenote: An issue bio.tools has with BioC
  All Bioconductor download links are invalid and/or broken

> [...] This is a known problem and its hard to convince Bioconductor people to keep old tarballs. What we do in Bioconda and Biocontainers is that we backup all used tarballs.

# Suggestions

- Add a https://www.bioconductor.org/sitemap.xml summarising site content to crawlers including google et al and TESS
- Migrate existing biocView Terms to EDAM / bio.tools ontology
- (Have) BioC packages imported to bio.tools on a regular basis (release? Weekly? Daily?)

---

Hi bio.tools team,

we've recently completed a review on >200 bioinformatics tools written in R for metabolomics data analysis, that we're now continuing to develop as a book [1].

Wouldn't it be cool if in the future a quick search on bio.tools [2] would get us those >200 packages ? Currently it is less than half, and we'd like to help getting that up.

One way I see this could improve is if we tell package authors how to best provide information that can be scraped by bio.tools [3]. For documentation, we could get a subset of your documentation [4] adapted to metabolomics and R into our book. Maybe some of the bio.tools team are even attending #EuroBioc2019 [5] and could initiate better data cross-talk between bio.tools and R/BioC ?
Just my thoughts on the train, Yours, Steffen

[1] https://rformassspectrometry.github.io/metaRbolomics-book/
[2] https://bio.tools/t?topic=Metabolomics&language=R
[3] https://github.com/bio-tools/biotoolsRegistry/issues/454
[4] https://biotools.readthedocs.io/en/latest/
[5] https://eurobioc2019.bioconductor.org/