Bioc Technical Advisory Board Minutes

3 November 2022

Attending: Vince Carey, Levi Waldron, Davide Risso, Lori Shepherd, Michael Love, Marcel Ramos, Charlotte Soneson, Stephanie Hicks, Aedin Culhane, Sean Davis, Kasper Hansen, Rafael Irizarry
Regrets: Wolfgang Huber, Robert Gentleman, Shila Ghazanfar, Laurent Gatto

:03 - :05 Approval of prior [minutes](#)

:05 - :19 SAB presentation outcome
- Pursue leadership in definition of interoperable representations for modern assays.
- Don't lose sight of reproducible research/open science objectives: new NIH data sharing requirements will be of concern and we can provide guidance/tooling. "Open science" concerns all research outputs, so that includes papers, but also data and software. In particular, open data is not possible without the associated software tools and algorithms.
- Bioconductor could be a leading group in the fields of data formats in single-cell/spatial.
- Take advantage of research in ML/AI.

:19 - :20 Outreachy/education progress [[sweave](#), BSgenome].

:20 - :25 Track funding sources in DESCRIPTION?
- Should we recommend developers to indicate funding sources in the Authors field in the DESCRIPTION file? Could be unwieldy if there are many authors.
- Funders probably like to see their portfolio of funded work (which isn't currently possible), and it would potentially encourage more funding of Bioc pkg development.
- Funding "fnd" may not represent that the funding was for a brief or finite period of time, but the same is true for "ctb" - we already don't put timelines to authorship or contributions listed in DESCRIPTION.
- Check CITATION files in packages - track citations by parsing the files and looking up citations.
- Proposal: track citation impact by having a flexible (multi-paper) format for the CITATION file (perhaps not ideal, but can work). It should be simple to do something like "what is the cumulative google scholar citations" to all "primary" Bioconductor papers. Should we also track funding in some other (non-DESCRIPTION) file then (depends on how that funding information may be used)?
- The Bioconductor *project* may want to report that software contributed *to* the project is funded by these grants.
- Goals:
  - Track citation output (for packages and project).
  - Track funding (tied to specific grants).

- Track key contributors to the project and enable communication (independent of funding for individual projects/packages).
- For the last point: Would be helpful if we could make sure that the same person uses the same name/email across packages.
- For the funding issue, is there any use in automatically parsing funding information from packages? I.e., does it add enough that it's worth pursuing?
- Standardizing names and citations would be highly beneficial to do soon.

:25 - :40 Ad hoc talks:
- Charlotte (license concepts/NumFocus requirements)
    - https://csoneson.github.io/bioc_pkg_licenses/bioc_pkg_licenses.html
    - OSI licenses: https://opensource.org/licenses/alphabetical
    - Another list of free software licenses: https://www.gnu.org/licenses/license-list.html
    - Bioc pkgs with only 'file LICENSE'
    - FAQ on 'non-commercial use'
    - Industry Working Group would like to write documentation about licensing to help inform the community.
- Aedin: slides

:40 - :50 Bioconductor 3.17 concepts
- Pause and establish definite priorities and road maps.
- Suggestions
    - BBS consuming too much effort? Prospects for renovation? Working group on actual requirements (cadence of building, synchronizing entire ecosystem, dealing with CRAN packages that come and go and come back).
    - Modularize the BBS
        - establish globally accessible database with package status recorded longitudinally
        - create build reports from database content
        - improve scalability of build process, containerize
    - Active maintainers must self-identify, edit DESCRIPTION or packages are deprecated
        - Maintainer must supply ORCID or explicitly opt out
        - Encouragement to have funding metadata in DESCRIPTION
    - GitHub Actions template should begin to take the place of single-package-builder
    - *Hub API renovation is a high priority
    - S4Arrays must progress
    - HDF5 in S3, HSDS, Zarr, TileDb back ends evaluated
    - BiocCheck
        - suggestion to require unit tests; coverage estimation is expensive, do in GitHub Actions? see slack discussion
        - ORCID
        - fnd (see above)

Comments:
- The build system is essential, but is very complex.
- Priorities:
    - feedback to individual developers (R CMD check/BiocCheck),
    - feedback at package submission
    - integrity of the entire project/interaction between packages
- Difference between feedback for longtime maintenance and feedback to developers who are trying to fix a specific issue

:50 - :51 Developer forum continuity (monthly on last Thursday? Mike and someone else).

:52 - :60 Other agenda elements
- liftOver project
    - Typical liftOver code has some manual steps - can probably make a wrapper to make that easier. A package that could do what is most commonly asked for.
    - AnnotationHub doesn't host chain files, just arranges for a user to access them - is that redistribution? We do automatically load them in R (but we could prevent that or require the file to be manually loaded). Get links directly from UCSC. Dispatch class is chain file - special treatment. File is cached locally (need to notify a user that the file has been cached locally). Can we get in touch with UCSC and ask to redistribute the files?
    - https://academic.oup.com/nargab/article/2/3/lqaa054/5881791
- interfacing with other major projects in the single cell or spatial field
    - Develop coordinated effort with software outside of Bioconductor. Interoperable classes, shared standards.
    - Packages like zellkonverter are especially useful for those working between computing environments.
    - Another option might be to generate joint containers?
    - Arrange a session at a conference (JSM, ENAR, …)? Or a joint meetup.

Tabled for future meeting:
- Discussion of common classes
- Suggestion for a "Glossies" working group. We need information about the project to inspire developers to engage and "contribute" their code, and information to present to sponsors about the benefits that come from sponsoring Bioconductor. We should have a clear menu of desirable services like careful independent review before acceptance, GitHub Actions template, regular build and check in release and devel (where enhancements can be made with limited restrictions), distribution on major platforms, support for containerization and cloud-oriented distribution in AnVIL, etc. ...