

# Bioc Technical Advisory Board Minutes

6 July 2023

**Present:** Charlotte Soneson, Michael Love, Vince Carey, Robert Shear, Lori Kern, Wolfgang Huber, Marcel Ramos, Stephanie Hicks, Aedin Culhane, Davide Risso, Kasper Hansen, Jennifer Wokaty, Alexandru Mahmoud, Shila Ghazanfar (from :30)

**Apologies:** Laurent Gatto, Levi Waldron, Sean Davis, Robert Gentleman, Rafael Irizarry

:00 - :03 Updates

- Previous [minutes](#) approved.
- TAB election processing is almost complete.
- August meeting is cancelled as it overlaps with BioC2023.

:03 - :18 Packages with no accessible maintainer.

- The core team has been working hard finding invalid emails that reduce our reputation for email services, eliminating them from our distribution lists. This raises the topics of 1) community adoption of packages that have no contactable maintainer and 2) deprecating and removing packages that have no contactable maintainer. TAB discussion and wording for guidelines will be welcome. 3) popular packages with a fair amount of reverse dependencies if no community members come forward (ex. aCGH?).
- [Bioc Orphaned Package Policy](#)
- Comments:
  - Can we ask authors to make a statement when submitting a package regarding what they would like to happen in the situation that they are no longer able to maintain it/if they can not be reached?
  - If someone finds the package useful enough to be willing to maintain it, that should be allowed.
  - Need to investigate licenses of older packages and see whether they would allow it; newly submitted packages must have an OSI-approved license.
  - Copyright theory connects to control over license
    - The recommended action for community members to take over an `unmaintained`/`unresponsive` package will be to keep any originally designated authors, move the current unresponsive maintainer as a c(`cph`, `aut`) designation for copyright holder and original author - and then add themselves as new `cre` maintainer.

:19 - :22 Concepts on collaborative authorship:

- Proposal in <https://hackmd.io/@mikelove/rJyu4OZF2> (editable [google doc](#))
- Encourage people to think about these questions in the early stages of a collaborative project.
- There are different situations:

- One person/group drives the project and invites others to contribute. This is perhaps the type that the current document is more directed towards.
- Collaborative projects - e.g. the OSCA paper. More difficult situation - more in need of guidelines? Speaks 'on behalf of the project' to a larger extent.
- Add some examples (good and 'bad' experiences).

:23 - :32 Mentoring process:

- The [discussion](#) on S4 pedigree is really great. Should there be some kind of live mentoring resource? We could have a 'BiocMentoring' repo and define an issue in which this conversation is conducted. Does that seem reasonable? Other significant discussions can occupy other issues. [Hervé replied that this particular discussion would be transferred to the issues of the Bioconductor classes working group].
- Engage with the mentorship program - there were resources created by the mentors/mentees of the first round.
- Discussion of S4 classes might fit best as a chapter in some kind of online living book / resource where "best practices for Bioconductor" are discussed. Package guidelines are already well described on the website, but maybe it's easier to keep a bookdown book.
  - <https://contributions.bioconductor.org/>
  - <https://github.com/mikelove/s4demo/>
  - <https://carpentries-incubator.github.io/bioc-project/05-s4.html>
  - revamp <http://contributions.bioconductor.org/r-code.html#class-design>

:33 - :40 Books:

- We will have to provide some guidance on submitting books to Bioconductor. This could be heavy and involve a working group and specs, or light with a single books repository that primarily links to content that authors stand up and stand behind. We could do some design of the repo or associated GitHub pages site, but this would be able to move forward quickly, provide systematic links that we could then include at [bioconductor.org](https://bioconductor.org), and we could evolve our approach to QA as time permits. Let's produce an action item.
- Different types of books:
  - single-author books
  - big collaborative efforts
- Needs a technical solution that is not overwhelming us.
- Things have changed since OSCA inception. Take advantage of things like GitHub Actions (need a good example, e.g. Hi-C book, microbiome book). May need a standard action for the books.
- Verification of scientific currency is important (but don't let perfect be the enemy of good, not realistic that everything is always cutting edge). Provide information of when the material was assessed for currency.
- There is a long tradition of having extensive documentation/vignettes in Bioconductor.
- Maria has started to engage with ELIXIR TeSS to see if we can create a Bioconductor TeSS for event and training <https://tess.elixir-europe.org/materials>. TeSS is also developing metrics for success of events and materials.

- Books can have more realistic data examples than vignettes - can also lead to problems with building. Take advantage of GitHub Actions (do they provide enough execution time/RAM?). We also have cloud storage and compute available. We may need something for capturing analysis output and using it as input to a book.
- Difference between books about a specific topic (from a single author) vs broader books (with more contributors, who may have different opinions on best practices).
- We need to take an 'editorial' role, showing the users best practices on how to use packages in combination, otherwise people will go elsewhere.
- One path forward may be to let the working groups handle this. E.g., if there is a microbiome working group and they want to write a book, then they can work towards it together, where others from Bioc are welcome to join if they wish. Then, if someone a year or two later feels like they want to change the book, they can join the working group and the group can discuss it and adopt the changes into the book.
- Having general guidelines that a book should at least contain a section on "Alternative points of view in this analysis field".
- Suggestion that books should be static entities released by editions fixed to Bioc release versions.
- A core aspect of our success is the between-package "competition" we have in our system. So caution is warranted about recommendations about best practices because that is the same as picking winners. At the same time, there are some established best practices/things to avoid.
- We can have multiple books on any topic. Do books get "published" / get DOI / citations?

:41 - :50 Python and machine learning

- Recent Bioc-devel discussion (e.g., <https://stat.ethz.ch/pipermail/bioc-devel/2023-July/019785.html>) on tensorflow support. Vince has a scikit-learn-dependent package (BiocSklearn) and a torch-dependent package (scviR) in Bioc, both using basilisk.
- BiocSklearn is throwing a strange error in devel, and scviR doesn't seem to port beyond intel linux. We don't have complete guidelines for interop with python -- 16.4 of [contributions.bioconductor.org](https://contributions.bioconductor.org) is terse and does not give examples, and does not deal with the issue of "cleaning" basilisk resources that become obsolete as package evolves.
- Posit provides [https://rstudio.github.io/reticulate/articles/python\\_dependencies.html](https://rstudio.github.io/reticulate/articles/python_dependencies.html) - is this usable in package configuration and installation?
- It seems that the amount of development of machine learning algorithms in python is far greater than that in R. [https://jax.readthedocs.io/en/latest/notebooks/autodiff\\_cookbook.html](https://jax.readthedocs.io/en/latest/notebooks/autodiff_cookbook.html) is one example of interest. Input from the TAB is welcome.
- Would it be good to have these discussions with JJ Allaire at BioC2023 to learn their directions and discuss basilisk?
- A minimal example of setting up a package with basilisk/keras would be helpful.

:51 - :60 Other topics

- Is there a Bioconductor container for single molecule footprinting data/is anyone working on something like this? The [SingleMoleculeFootprinting](#) package is using a standard matrix.
- `sp` package largely superseded by sf and terra, but at least a few Bioc packages are importing sp <https://r-spatial.org/r/2023/05/15/evolution4.html>.
- <https://www.bds3.org/> is up and running and going well so far. With Laurent, Robert, Helena, Wolfgang (<https://www.bds3.org/faculty>).