Bioc Technical Advisory Board Minutes

6 October 2022

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

:03 - :04 Previous minutes approved.

:04 - :06 CZI meeting – nice discussion with Zarr community
- representation of large numeric arrays
- new version of the specs coming out

:06 - :08 Joint CAB/TAB meeting comments
- How much do the different boards know about how the other board's function/what they do? Shared understanding of the objectives of each board, how do we say whether the boards are successful at achieving these objectives?

:08 - :30 Ad hoc talks:
- Shila (on submitting raw seqFISH imaging data to IDR)
    - Is there a good resource for newcomers? OSTA book (is there a slack channel?)
    - Why BumpyMatrix? Capturing not just gene abundance per cell, but molecules with coordinates. Genes x cells x "extra" (bumpy matrix bit).
    - IDR Github links - OME, rOMERO
- Alex
    - 10 google nodes used each Monday to build binaries. Use Jetstream instead? Or GitHub Actions: https://github.com/almahmoud/gha-build. Having these binaries is crucial e.g. on AnVIL.

:30 - :47 Upcoming Scientific Advisory Board Meeting. Develop a charter for the SAB to guide them - think about what kind of advice is useful to get, current pain points.
- What do the SAB members consider essential in order to start/continue using Bioconductor?
- Should we pursue the books more systematically (as additional products from the project)? Also discuss how to best do this in practice.
- Format of the meeting? Do the SAB members first get an update on the project? In the past they were given the annual report - Vince is working on a report, aimed to be finalized next week.
- Consider including appendices of activities of TAB/CAB. For example, the TAB lightning talks give a really nice taste of some of the cutting edge work being done across the

project. Other fodder includes blog posts, papers published, etc. All in appendices, though.
- How much money/resources should we spend on e.g. making data sets available to the community, having a shiny server (see below), etc?
- Highlight early in advance what the SAB should read in order to prepare.
- Need to have a mission statement/objectives to get good input on what to do to reach those goals.
- Limited amount of time for the meeting - be concrete of what you want to get to.
- Do SABs usually have a charter? Makes it easier for the SAB - outline what the project wants from the SAB.
- Sometimes we get 'boxed in' because of history - bring up some of the points that have been recently discussed in TAB/CAB with the SAB. E.g. cloud usage, inclusivity & diversity.
- Is there usually a limited number of topics that the SAB is tasked with for a given meeting? How to prioritize? Put the important things first, and share the list in advance.
- Create a document with possible topics for discussion in the TAB/CAB meeting next week.
- Are we trying to get people to use the cloud more - why? How does it relate to the goal of Bioconductor? How do we achieve that?
- How do we help new developers?
- How do we get 'big' projects/more developers to be part of Bioconductor?
- How to get more users - onboarding?
- SAB Success Tips (from Robert)
    - Project a Mission Statement. Clear mission (what you want to achieve). Clear goals & objectives.
    - Meeting should define the project. Here is what we are doing. Will they be effective or are they important.
    - SAB can advise on goals that will be productive.
    - SAB can suggest ideas or discuss barriers to achievement.
    - Best results if the questions to the SAB are direct and specific.

:47 - :52 Shiny – should foundation/grant pay for significant shinyapps.io allocation? Could also do something in Jetstream2.
- BiocViews: Shiny
- EuroBioc Shiny BoF - may develop as an article / white paper. Already expanded in contributors guide.
- Costs/benefits of Bioc supporting a shiny deployment platform
    - Could be a highly useful & popular service, e.g. given experience at EMBL - but is it our mission? Collaborate with academic cloud providers (e.g. ELIXIR in Europe)?
    - For how long is that commitment? ∞? Life cycle / migration out policy? Could depend on use/visits.
- What would be considered a success if Bioc did invest?

- What are the personnel costs?
- What would the Bioconductor core team's involvement be?
- Is shiny a cost-effective approach for deploying apps that become "popular"?
- What should it be used for? One idea is as a 'gallery' of Bioc shiny apps, a bit like interactive vignettes.
- Not Bioconductor's mission to provide shiny apps for analysing data - hard to manage cost for compute.
- Having a showcase gallery could be a good idea.

:52 - :60 SparseArray preview (see also presentation from BioC2022 on YouTube)

- sparse data representation in R
- current de facto standard: dgCMatrix
    - limitations: only for 2D data, no more than 2^31 nonzero values, nonzero values always stored as doubles, some operations are inefficient
- matrixStats (CRAN)/sparseMatrixStats (Bioc)
- SparseArray objects (implemented in SparseArray package).
    - support an arbitrary number of dimensions
    - support any R atomic type
    - support an unlimited number of nonzero values (but individual dimensions must be <2^31)

Open discussion

- Large data/file distribution best practices (options, documentation) -> postpone until next meeting
    - https://contributions.bioconductor.org/data.html#data
    - Can we identify other cloud (with free/reduced egress) or write an application to use DICE etc to host large data sources?