

Bioc Technical Advisory Board Minutes

4 November 2021

Attending: Vince Carey, Michael Love, Lori Shepherd, Laurent Gatto, Charlotte Soneson, Hervé Pagès, Levi Waldron, Nitesh Turaga, Shila Ghazanfar, Marcel Ramos, Rafael Irizarry, Wolfgang Huber (until :20), Jennifer Wokaty, Kasper Hansen, Aedin Culhane, Stephanie Hicks
Regrets: Robert Gentleman, Aaron Lun, Hector Corrada Bravo

:03 - :04 Previous [minutes](#) approved

:05 - :15 EuroBioC 2022 planning discussion

- The [website](#) is currently being edited/finalized
- TAB satellite meeting - before or during the conference?
- Presentation about current core activities (e.g. Azure, hubs, build system replication locally)
- Aim to get more people engaged (submit packages, review packages, ...)
- First Bioc Carpentries workshop on Mon/Tue in conference week (independent of conference, but advertise via conference channels/with the conference to promote the attendance of newcomers)
- Session to introduce existing and non-active but needed working groups
- Developer day rebranding -> Collaboration day, Community day, Development day, Software day, Training Day
- Bring in the mentorship program? (Lots of things happening there at the moment)

:16 - :20 Greatest hits

- Release 3.14 out, no major crises, great job by core
- Martin Morgan spearheaded coordination of CZI Single Cell initiative. This was mentioned to John Readey of the HDF Group. HDF5 is a critical component of current solutions, and it would be nice to share some support with the HDF group. The data structure work is run by Davide and a foreign grantee can't have a subcontract back to USA. [Brief notes](#). Mike Smith is working on adapting rhdf5 to single-cell context etc. Reach out to the scanpy/AnnData developers?
- Two core members giving keynotes at BiocAsia: Lori Kern (What is Bioconductor; overview of core activities), Nitesh Turaga (Efficient use of Bioconductor in cloud computing environment); Nitesh also spoke at the Turkish Biostatistics conference. Workshops at BiocAsia by Aedin Culhane, Federico Marini, Laurent Gatto & Christophe Vanderaa
- ELIXIR [bio.tools](#) - would the underlying API be useful for fostering discovery of Bioc resources? Does it need enhancement?

:21 - :30 CAB topics

- The last CAB meeting was joint with TAB - approve [minutes](#)

- Updated BiocAwards mechanisms, approved new Mission Statement, new working group on import/export classes (Vince, Laurent, Jo), Rladies Liason (Susan and Daniela), GDPR progress (contact with Stephanie Butland).
- [BioC Asia 2021](#), Nov 1-4 2021. Keynotes and workshops (mentioned above)
- H3ACC and H3ABioNet workshops (Nov 2021). Identified 4 Workshops from BioC2021:
 - Introduction to Bioconductor annotation resources / public data resources
 - Dimension Reduction for Beginners
 - systemPipe: Workflow and Visualization Toolkit ([systemPipeWorkshop2021](#) • [systemPipeWorkshop2021](#))
 - Multi-omic Integration of cBioPortal and TCGA data with MultiAssayExperiment
- Translation (Lori's talk with Japanese subtitles: <https://bit.ly/WhatisBioconductor>)
- Mentorship program: Met with new mentors, slack channel, establish ground rules, open recruitment for mentees shortly (up to 12)
- BiocAfrica event? Bring together/coordinate groups active in different parts of Africa.

:31 - :35 TAB Working groups touch base - roll call, add links here, note [BiocChallenges](#)

- Open problems https://openproblems.bio/neurips_2021/ will be presenting at [NeurIPs](#); registration is open. There is incentive if ties to meeting and publication.
- Hubs: following conversation from BioC2021, Mikhail Dozmorov has added [excluderanges](#) to AHub in 3.14, which provides access and metadata on sets like ENCODE exclude list.

:36 - :45 TAB discussion of "scope of service" of the Bioconductor core team and project. We don't formally define our services and often expand our activities to solve problems or add new facilities.

- Daily builds with full checks for devel branch, developers rely on outcomes and often cannot reproduce -> containers
 - Transition to microsoft
 - UCRT new Windows toolchain (help with multilingual data representation)
 - Can "incremental" building (on change, tracing dependencies) save resources and allow more frequent feedback to contributors? Encourage developers to test locally (but need to be able to reproduce the settings of the build system/have access to all platforms - can we provide temporary access to the SPB?)
 - Less frequent builds for release, other products
- Global content delivery via CloudFront - what alternatives are there? Thanks Wolfgang for doing some research: Alternatives to explore
 - Fastly and open source program
 - European Open Science Cloud
 - EMBL with EBI/NCBI/DDBJ
 - RStudio CDN (0-Cloud for CRAN)
 - <https://stackoverflow.com/questions/16875174/what-does-the-0-cloud-mirror-stands-for-in-r>
 - <https://blog.cloudflare.com/introducing-r2-object-storage/> - AWS contacted by Nitesh, acknowledgment received but no details on possibilities.

- Mirrors? Need to understand how they interact with BiocManager. Compared to CRAN, Bioc is hosting lots of data - fewer packages, larger total size.
- Ad hoc rsync services
- Hubs and hub API, transition to microsoft
 - large topic; NCI ITCR-oriented discussion with Nathan Sheffield
- Package reviews, git management, SPB
- [rising in importance] Distribution of built binaries for Linux
- [significant] Creation of binaries for M1 mac
- Legacy packages (Bioconductor 3.1 and earlier) can be moved to Azure
- Big question: can the core developers reduce their maintenance efforts (pivoting to improved QC of internal stacks and advanced development) but keep general service satisfaction constant or increasing, by improving distributed maintenance and delivery?
- What is the role of GitHub actions for advanced contributors? Can its use by contributing developers lead to reduced QC interventions by core?
 - Once you start using GHA for package check, you get the nice pkgdown with little extra effort, better navigation than Bioc's landing page across vignettes/function man pages
- What is the role of containerization for developers on the TAB? How do you use containers in package maintenance?
- How can we simplify git self-management?

:46 - :55 Communications and mobilization of community

- Role of pkgdown -- seems effectively used by tidyverse and tidymodels to invite exploration. Should we produce a community agenda or hire a student to do the conversions of Sweave that are a first step to achieving pkgdown ability?
 - Generally, we benefit by choosing a system and providing some documentation/standardization/support for that.
 - Can we have cross-package links between man pages?
 - GitHub repo main branch corresponds to Bioc devel branch - may not correspond to what you get in the Bioc release branch
 - biocthis takes care of a lot of the issues
 - Manual linking to compiled pdf vignettes (in yaml header)?
 - New packages, or all packages (svn, Rnw vignettes, no GitHub repo, ...)?
 - Discussion regarding pkgdown alternatives on [Slack](#)
- BiocChallenges, and some aspects of renewal of project elements - Rsamtools examples based on hg18, Sweave conversion, ...
- Project health dashboard, pkgnet, test coverage

:56 - :60 Open floor -- Thoughts about boundaries to cross to declare Bioconductor 4.0 achieved, and goals for 4.x

- Increasing diversity of developer pool (consistent with NHGRI [Forefront](#) goal)
- Improving educational outreach to US underrepresented groups (CZI D&I as a step towards HS and undergrad contact)

- Simplifying access to genomics on a broader ancestral base (AllOfUs, Million veterans, ...)
- Concerted efforts at de-siloing computational genomics, e.g., special attention to cross-language interfaces, API-focused designs, microservices
- Moving the builders to a different location
- Transition of BBS to using containers?