

Bioc Technical Advisory Board Minutes

5 January 2023

Attending: Vince Carey, Levi Waldron, Laurent Gatto, Aedin Culhane, Alexandru Mahmoud, Lori Shepherd, Charlotte Soneson, Rafael Irizarry, Jennifer Wokaty, Robert Gentleman, Kasper Hansen, Marcel Ramos, Stephanie Hicks

Regrets: Shila Ghazanfar, Mike Love, Sean Davis, Davide Risso, Wolfgang Huber

:03 - :05 Prior [minutes](#) approved.

:05 - :10 Update on Carpentries instructor training:

- 14 applicants accepted and signed up for instructor training (Carpentries membership gives us 15 seats per year). From US, Europe, Australia, Asia, Africa, covering multiple languages. Checkout process ongoing for those attending instructor training before Christmas.
- Aim to arrange workshops preceding BioC2023 and EuroBioC2023.
- The [bioc-intro](#) lesson is in beta phase - needs 'external' trainers to run the workshop and provide feedback before it can be officially included as a Carpentries lesson.
- Any contributions are welcome - a list of the currently active lessons can be found at <https://bioconductor.org/help/education-training/>.

:10 - :20 "Bioconductor Project" accepted for fiscal sponsorship by NumFocus.

- Motivation: growth and complexity of Bioc Foundation.
- Acceptance of fiscal sponsorship agreement (FSA).
 - Bioconductor foundation assets will be transitioned.
 - Services to be provided by NumFocus include fundraising organization, banking services, asset services, admin services, limited liability protection, legal services, marketing services, negotiation of contracts, and disbursement of project funds.
- We need to work out the human interface to NumFocus administration (Maria Doyle primary, another backup will be needed).
 - Executive committees of TAB and CAB will have meetings with NumFocus in early 2023 to establish general operating approaches.
- Bioc conference planning.
- Solicitation of sponsors and unrestricted donations.
- Establishment of grant application/review/assessment procedures. This has nothing to do with NIH grants at this time, but concerns "internal" grants to projects related to Bioconductor that are outside the scope of NIH or other institutional support. Safest approach is to grant reimbursement for documented pre-approved expenditures. Would work for ad hoc conference attendance, travel, small hardware purchases. For larger projects a budget with specifications of needs and deliverables and a fractional payment approach.

:20 - :25 How do we maintain a leading role in a world that is moving to multimodal and image-oriented genomics, where more dev types are strongly connected to python/conda?

- Are incoming grad students and postdocs better versed in R or python?
- Is development of R packages a skill of recognized importance, or are scripts in GitHub considered sufficient for research?
- Spatial transcriptomics/image analysis methods often in python (including DL frameworks). On the Bioc side, infrastructure is needed. Currently, lots of transfer of data between different formats.
- Similar for proteomics/metabolomics R and python ecosystems.
- Bioconductor well respected in the statistical data analysis realm - complementary to much of what is the focus in python. Python is stronger/directly supported on the deep learning side. Make it easy to move between frameworks.
- Lots of packages on GitHub depend on underlying packages from CRAN/Bioconductor, but are not contributed back to the projects. Do we have a good enough description of the benefits of contributing?
- One noticed barrier to making data analyses more interoperable is that people who *want* to do that, often don't know *how* to do it. So it is easy to form the opinion that "it's not possible to do in R/Bioconductor". Develop training materials to demonstrate how to do this.

:25 - :32 Evaluate the [r2u discipline](#) for feasibility as a package manager-driven provider of Bioc software for container users (or those using the supported flavors of Linux).

- How valuable would this be for TAB members? How many are using containers routinely? For the community?
- The reason Vince started looking at this was the need to establish an R-enabled Spark cluster in terra.bio, for [BiocHail](#). He wound up making terra *build R from source* as a step in defining the cloud environment for the cluster.
- Will you get the latest version of packages (general question with package managers)?
- Could it be used to build Windows/mac binaries? Windows binaries are probably well covered at the moment, mac less clear. What about Linux ARM?
- Current Bioc Linux binaries are not going away, work well in containers that support them.

:32 - :35 Bioc 3.17 release by normal means, except a) want default devel branch name to be "devel", not "master" and b) macOS ARM builds for both release and devel.

- Marcel's package ReleaseLaunch and [renaming utility](#).
- Could the branch be checked via BiocCheck? May not be straightforward - branch names do not have to be the same locally/on Bioc git. BiocCheck is currently not routinely run on all packages (only at submission).

:35 - :38 Better planning for presence at large conferences: ISMB, useR, JSM, ENAR, [ABRE](#), RECOMB.

- Who goes, who presents items related to Bioconductor?

- Two tutorial submissions for ISMB 2023:
 - Davide
 - Vince ([last-minute](#) on 'ecosystems for genomic data science')
- How can we help people coordinate attendance/presentation/...? (also being discussed in the CAB)

:38 - :40 Better planning for significant publications:

- NAR for Hubs (database issue deadline in August, but need to contact the editor before July 1st. https://academic.oup.com/nar/pages/Ms_Prep_Database).
- R Journal pieces for each issue (editor solicits "news from Bioconductor", but needs more).

:40 - :57 Developer forum

- Third Thursday, January - possibly Marcel on ReleaseLaunch, additional volunteers.
- Core devs should use that date to present current work and opportunities for contributions from tech-savvy developers.
- "Endorsed classes" working group could appear.
- Working group concept overview.
- Other potential topics:
 - Interoperability
 - reticulate, basilisk to deal with python – maintenance of basilisk and dealing with exotic or outdated systems, configuration with conda.
 - Teamwork on basilisk to ensure that it works on standard systems.
 - Data structures
 - AnnData, HDF5 backend.
 - TileDb.
 - Zarr.
 - Continuous integration and GHA
 - rworkflows (Brian Schilder).
 - biothis.
 - CI for the build system could be GHA.
 - GHA for binary repo production is already working well: <https://github.com/almahmoud/gha-build>.
 - What happens if GitHub Actions goes away? Could we build something that could work on another platform as well? For the binaries, the first version was run on Kubernetes/Jetstream, the current setup can be relatively easily moved from GHA if needed. Harder for Windows/mac (i.e., if we replace the entire build system), as we would need the runners.
 - Suggestion to reach out to other organizations maintaining/distributing large numbers of R packages (e.g. Neuroconductor. ROpenSci, PharmaVerse).
 - We have a substantial allocation from NSF for compute on Jetstream2. Also have storage allocation.

:57 - :60 Website discussion:

- Are static websites "a thing of the past"? A personally configurable web resource might be interesting - are we looking into new resources?
- Developer resources, build reports etc are ideally suited for a static website.
- The goal is to provide an easier on-ramp for people.
- The website needs to cater to many different groups of users (beginners as well as experienced developers).
- Providing 'links to everything' makes it hard to find things (some things are repeated in multiple places on the landing page).

Other topics (not discussed):

- Formats for future meetings - breakout rooms for the first 15 minutes (topically organized, topics provided by working groups)?
- Consider a traveling monthly event in various cities with a topic of interest (technical focus) and hands-on workshop.
- Examine the collection of contributed packages for scope of scientific impact.
- Increase visibility of key processes: how many cloud machines, how much storage, how fast and frequently do we build, are all Annotation/ExperimentHub entries valid, are mirrors complete and consistent, etc.
 - 300TB now OSN approved, S3-compatible.
- Enhance visibility and usability of containers for end-users. Our binary repos save a lot of compilation time, r2u and debian approaches save even more.
- Technical gaps that TAB members are encountering.
- Assessment of contributions: are scientific domains being missed? Do we need to change our approach to providing annotation, data, software components so that methods leaders can take advantage of them and join the ecosystem?
- ARM platforms (text below from <https://jumpcloud.com/blog/why-should-you-use-arm64>):
 - *For example, in November 2020, Apple Inc. unveiled the first Macs with ARM64-based M1 chips, debuting new MacBook Air, MacBook Pro, and Mac mini models. The company also announced plans to fully transition away from Intel to in-house chips over the next couple years. In June 2021, Microsoft Corporation announced that it is building native and interoperable applications for Windows 11 on ARM64 processors. Both Microsoft Corporation and Apple Inc. are leaning into the low-power consumption and enhanced computational efficiencies in ARM64 chips to reverse a steady decline in PC and Mac sales. In 2018, Amazon Web Services (AWS) unveiled Graviton processors—64-bit ARM chips—to power its Linux-based Amazon Elastic Compute Cloud (Amazon EC2) instances. AWS Graviton processors also support popular Linux operating systems such as Amazon Linux 2, Red Hat, and Ubuntu. AWS has since enhanced the capabilities of Graviton processors to deliver optimal performance, particularly in EC2 instances. The growth of Graviton processing in AWS has been driven by both significant cost reduction and strong processing power, both characteristics that are good for business.*