Bioc Technical Advisory Board Minutes

1 May 2025

Present: Vince Carey, Stephanie Hicks, Wolfgang Huber, Michael Lawrence, Marcel Ramos, Erdal Cosgun, Henrik Bengtsson, Ludwig Geistlinger, Rafael Irizarry, Lori Kern, Hervé Pagès, Laurent Gatto, Andres Wokaty, Charlotte Soneson, Alexandru Mahmoud, Levi Waldron, Kasper Hansen
Apologies: Jacques Serizay, Dario Righelli, Helena Crowell, Davide Risso

:03 - :04 Note that NHGRI has restored some of the funding cut for year 5 of the current grant; next competitive renewal will be reviewed in October.

:04 - :08 Slack pro for community must be shut down this month.
- See here for some info about exported content.
- Options: discourse, unify support, slack channeling, bioc-devel mailing list, zulip, discord.
- Must move bioc-devel off ethz hosting, Linux Foundation may pick it up.
  - groups.io, mailman hosting, affordable rate per year
  - bioc-devel will move autonomously
- Could GNE fund slack through a grant? What would be the benefits to GNE?
  - Knowledgebase for mining, training, dedicated recruiting channel, "sponsored by…" banner

:08 - :21 Approve minutes, overview of working groups
- classes
  - Note new Seqinfo package to reduce burden on GenomeInfoDb.
- cloud
- EDAM/biocViews
- mass spec
- histopath image analysis
- package reviews
- social media and blogging
  - R consortium is involved in blogging/interviewing - professionalization. Could CAB take this up?
  - Rticles, mastodon/bluesky, LinkedIn.
  - Should we go back to posting on X?
    - Lori to bring discussion to CAB.
    - Passively autoposting bot - no engagement
    - https://blog.bioconductor.org/posts/2023-11-17-twitter-exit/
- spatial omics
- workflows
- multilingual working group (French translation subgroup):

- ○ bioc-intro has been translated to French. Course not rendered yet, but should become available at https://bioconductor-translations.github.io/bioc-intro-fr/
- we need entries in the workinggroups document on pangenome and wasm
  - ○ pangenome:
    - ■ Tim mentioned FindMyFriends which was dropped in Bioc 3.13; vignette produces various http errors and uses obsolete dplyr functions; should package be adopted?
    - ■ Tim mentioned GFAtools, and privacy-preserving sequence simulation
    - ■ Tim suggested GA4K; should we discuss possibility of a "curated" package, possibly "data-free" for those with access
    - ■ minimap2 interface provided by Shian Su - see some questions here.
    - ■ See slide 24 at previous meeting's deck; is an R03 suitable?
  - ○ wasm
    - ■ In-house compilations in progress
    - ■ Experimentation with new packages is not well-supported at this time. By this I mean that once you compile to emscripten using an action or container, to run your package in local webR session for testing involves various security-weakening steps in Chrome.
    - ■ See this repo, which produces wasm binary packages for a munged version of SummarizedExperiment.
    - ■ r-universe ships emscripten artifacts that don't work because of unmet dependencies; suggestion on mitigation has been offered by Jeroen (make reliance on annotation packages "soft", don't import).

:21 - :37 Release binaries
- Binaries for several (also core) packages not yet available in release for macOS silicon - is there a systematic reason, is anything additional needed on the hardware side?
- Could set up automatic notifications also for other platforms than Linux.
- There's a bit of delay for the mac builders because the builders have to be switched.
- Is the build system up to date? Yes. If so, it is up to the developer to solve absence of binary.
- Should "core packages" be tested and required to be available before a release?
  - ○ Which are these "rate-limiting" core packages?
  - ○ Community engagement on definition of core packages
  - ○ "Rapid builds" on a set of core/infrastructure packages (updated every 6 hours): https://bioconductor.org/checkResults/3.22/bioc-rapid-LATEST/
  - ○ Visibility of events of this type
  - ○ Should these packages be required to build on all platforms? Most?

:37 - :38 Overview of R-universe for Bioconductor
- Get familar with bioc.r-universe.dev
- We need to replace the git "hooks" that control developer changes to central code image (e.g., version numbering sequence is increasing with changes), this will consist of "gatekeeper actions".

- Would we save money if we moved everything to r-universe today? Possibly on hardware, reduction in egress cost for AWS. Would reduce humanpower needed to maintain builders.

:38 - :42 Technical steps forward.
- Adaptation of BiocStyle to work with quarto - [GitHub issue](#) with discussion.
- GPU-dependent packages [can be tested](#).
- Proposed modification to basilisk, see [rationale](#) and currently open [PR](#).

:42 - :60 Discussion of a sustainability working group.
- This group is needed to establish governance "beyond USA".
- Should have two components: "funding" with members capable of applying for and administering new funds, and "requirements", with members articulating needs for project development.
- A general concern for new research projects: include "subcontracting" to core developers and DFCI and NumFOCUS to cover expenses of Bioc personnel and ecosystem operation/maintenance.
- At risk in case of severe NIH cuts are the review/ingestion processes, distribution via CDN (the transition to r-universe might mitigate this), update and maintenance processes for core software and annotation, container/binary production, coordinated releases with evolution of R, communication and support services like support site, mailing lists, probably more. We've never put price tags on these nor estimated the economic value to users.
- UK Research Software Maintenance Fund: https://www.software.ac.uk/research-software-maintenance-fund/round-1 (requires a project lead in the UK, letter of intent 2025-05-30).
- ELIXIR.
- Several groups in Europe are hiring developers, with the idea that these should also be able to contribute to core activities.
- Main bottleneck may be to split up the work - more management involved if work is divided, but would give more resilience. "Formal" project management.
- What can be done? Virtualize the build system so that copies or clones could run elsewhere? Consider core team members employed by European institutions that are co-supervised by core team lead.
- How much technical debt do we have? Are we held back because of this?
  - Annotation system is using old technologies, hasn't been revised recently, used by many, pervasive in the ecosystem.
  - Move towards S7 to get rid of reliance on the methods package.
  - AWS infrastructure and operating systems are up to date.
- Sparse matrix representation - size limitation with Matrix package, SparseArray addresses that. Should/will there be something more general/in R itself? Would likely be a lot of work to update the Matrix package (large C code base). Could be a topic for an (ISC) working group, beyond Bioconductor.