

Bioc Technical Advisory Board Minutes

2 November 2023

Present: Vince Carey, Levi Waldron, Charlotte Soneson, Lori Kern, Helena Crowell, Mike Smith, Ludwig Geistlinger, Jen Wokaty, Alexandru Mahmoud, Hervé Pagès, Brian Schilder, Davide Risso, Rafael Irizarry, Stephanie Hicks, Marcel Ramos, Sean Davis (from :40), Henrik Bengtsson (from :45), Wolfgang Huber (from :50)

Apologies: Michael Love, Maria Doyle, Laurent Gatto, Kasper Hansen

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

:04 - :09 Updates

- Bioconductor 3.18 released.
- Much progress on [bioc2u](#) and engagement with Iñaki Ucar on *self-contained* binary (e.g, .deb, .rpm) generation for other linux flavors.
- [sysreqs](#).
- AWS modernization progress.

:09 - :10 Accepted LOIs for CZI EOSS-6 include

- OSCA (Ludwig, Davide, Stephanie, Pete)
- iSEE for spatial omics (Kevin, Federico, Charlotte, Ludwig)
- GPUs (Levi, Davide, Gabriele, Vince)
- democratize training platform (Alex, Enis, Nuwan)
- ontological resource tagging and discovery (Vince)
- Zarr for R and Bioconductor (Wolfgang & Mike S)
- NetworkHub (Federico, Sebastian, Vince)
- R for mass spectrometry (Laurent, Johannes)
- Multi-omics analysis (Himmel)
- Tidyomics (Mike, Stefano)
- Supporting and sustaining Bioconductor developers (Maria, Aedin)
- Delivering Bioconductor Training in Africa & elsewhere (Aedin, Maria, Laurent, Charlotte)

:10 - :12 Sean's [query on slack](#) concerning team support for packages/multiple maintainers.

- Commit rights as currency for open source, dispute resolution.
- Note that the avoidance of conflicts is an important justification of restricting push rights to a single individual.
- Would require changes to R CMD check (does not allow multiple maintainers).

:12 - :20 Followup on build system presentations from last call.

- BBS container for Linux replicated in `ghcr.io/bioconductor/bioconductor_salt:jammy-r-4.3.1`.
- Rtools43 updated ARM 64 on Windows.

- Mac Pro arrived and is currently being configured.
- Learn about [meson](#) and related multiplatform build systems (Levi reached out to meson developers). NumPy is using meson.
- Easy to use python from R, harder to use R from python (outdated bioconda packages).

:20 - :22 Expanding the build system to incorporate non-R assets? Collaborating with external redistributors.

- bioconda: there seem to be issues implying that RcppParallel and RcppEigen won't build when the r-4.3 from bioconda is used:

```
I used a jetstream2 instance with ubuntu 22.04. I ran
conda install -c conda-forge r-base=4.3.1

. I then installed BiocManager with install.packages and attempted
BiocManager::install("dada2"). This ran afoul of
checking whether the C++ compiler supports the long long type... no

*** *****
*** stringi cannot be built with these settings in place.
*** See the INSTALL file for the solutions to the most common problems.
*** Moreover, explore the list of open and closed issues at
*** https://github.com/gagolews/stringi/issues/
```

- [macports](#) are also distributing Bioconductor packages. Both Bioconductor and CRAN have very specific approaches to distributing software packages for macs. When the documented methods are followed, it seems straightforward to obtain consistent installations using BiocManager. <https://ports.macports.org/port/R-GenomicRanges/builds/> is on rtracklayer 1.55.0, which would not be a version we'd propose using with current R-patched.
- Software redistribution through, e.g., macports or conda leads to complexities that might be reduced with more collaborative communication.
- A new channel #other-distros has been created at the Bioconductor community slack so we can have a more inclusive discussion if such is of interest.
- https://en.opensuse.org/Portal:Build_Service

:22 - :60 Helena presentation on spatial transcriptomics infrastructure

- [Slides](#)
- SpatialData: <https://www.biorxiv.org/content/10.1101/2023.05.05.539647v1>
- Layers: tables, points, shapes, labels, images.
- Feature aggregation (spatial queries), coordinate system alignment, multi-modal visualization.
- Corresponding R package under development.
- ZarrArray - no delayed backend yet (proposed in Mike's EOSS grant application - but this will be some time into the future).
- Images can be aligned (on the fly).
- SpatialDataIO - readers for platforms with standardized file formats.
- SpatialDataUtils - maskings, e.g., images by labels/shapes, labels by shapes, points by labels/shapes. Cropping according to annotation.
- SpatialDataPlots - transformations, element-wise aesthetics control.

- SpatialExperiment not capable of representing all these layers - can pull out a SPE from a SpatialData object.
- SpatialData works directly with Zarr/parquet.
- Preprocessing/image processing could be done outside of Bioconductor, and read into R as a SE-like object?
- Major current roadblocks? Images, large molecule data frames - currently don't fit into memory in any of the existing Bioconductor classes (unless working on cell-level summaries).
- Image analysis will likely not be done purely in R (except visualization).
- However, like for any data analysis, there is a loop that includes visualization, statistical downstream analysis, quality assessment of whether the upstream was good enough, and so on - there is a role for R here.
- Working with the polygons etc in R/Bioconductor would be nice.
- geoparquet seems to be efficient/working well in R.
- What about leveraging/wrapping the existing python implementation? Currently not stable - having an R implementation would make us independent.
- https://bioconductor.org/packages/release/bioc/vignettes/DelayedArray/inst/doc/02-Implementing_a_backend.html
- We don't necessarily have to use Zarr for everything - parquet, TileDb, HDF5 (as an interim solution?). But Zarr is commonly used in the (bio)imaging field. Giotto uses [terra](#). Zarr is more cloud-compatible than HDF5.
- Spatial transcriptomics working group - bring interested people together.
- Think about the intermediate steps as well (example data, documentation, benchmarks, ...).

Some older business reviewed.

- Vince reached out to CRAN to ask when Quarto will be supported for vignettes - not clear yet. Henrik noted in minutes: The quarto package does not support Quarto vignettes, but that's easy to add. Indeed, there is a PR (<https://github.com/quarto-dev/quarto-r/pull/57>); verified with a standalone package (<https://github.com/quarto-dev/quarto-r/pull/57#issuecomment-1751093594>). There's also a discussion at <https://github.com/quarto-dev/quarto-cli/discussions/2307>.
 - **News Oct 24:** via CRAN contacts, posit will create engines; CRAN has installed the quarto cli support. "Work in progress".