Bioc Technical Advisory Board Minutes

2 May 2024

Present: Henrik Bengtsson, Vince Carey, Levi Waldron, Laurent Gatto, Ludwig Geistlinger, Charlotte Soneson, Michael Love, Marcel Ramos, Hervé Pagès, Lori Kern, Robert Shear, Erdal Cosgun, Alexandru Mahmoud, Davide Risso, Jen Wokaty, Wolfgang Huber, Maria Doyle, Sean Davis
Apologies: Rafael Irizarry, Stephanie Hicks, Helena Crowell, Kasper Hansen

:03 - :05 Previous minutes approved.

:05 - :08 Bioc Awards nominations end May 15, look for Google Form link on this page.

:08 - :18 Cloud Methods Working group update
- Strategic plan: Draft v.0.1, feedback welcome.
- LLM for Bioconductor discussion started.
- Observation in renewal for NHGRI: they encourage taking advantage of AnVIL/Terra development. Have to list plans for data management/sharing.
    - From AnVIL, there is a Cloud Cost Budget Justification template: https://anvilproject.org/learn/investigators/budget-templates.
    - AnVIL for DMS plans (ASHG slides):
      📄 AnVIL-for-DMS-ASHG-09.14.2023_final
    - Google Pricing Tables (needed for up-to-date budgeting): https://cloud.google.com/storage/pricing
- Also in Europe there are pushes towards using cloud resources, e.g. European Open Science Cloud.
- Don't focus just on the compute - it's also a collaboration environment. Safe data storage, controlled access. Useful for inter-institutional collaborations.
- One of the ideas of AnVIL is that you should do your computing there and avoid egress.
- Cloudflare R2 (S3 compatible) for egress-free data sharing. https://www.cloudflare.com/developer-platform/r2/
- Document where cloud computing/storage is currently used in the Bioconductor community/ecosystem.
- Usage for educational purposes.
- Bioconductor should help people take advantage of the cloud where possible.
- NIH has made some progress in making cloud available and teaching people to use it: https://cloud.nih.gov/resources/cloudlab. $500 free credits and 90 days of usage of a cloud, with self-directed learning modules.
    - Can enable more equitable computational biology.
    - Not being used for R01s yet (apparently).

- Think about cloud as an institutional resource. Cost sharing. The same logic used for on-premise resources like HPC can apply to the cloud (the business model is well understood).

:18 - :24 Zarr and spatial
- Passage to Zarr v3 spec
- Think of Rhdf5lib bundling a compressed C++ library to avoid any middleware
- The following is found at the [Zarr Enhancement Proposal](#) repo
  > *Interoperability. Zarr v2 has been implemented in Python, C, C++, Julia, Java and JavaScript. However, there is not feature parity across all implementations. Implementations are also not currently available for some important languages, such as R. This is in part because the Zarr v2 spec was originally developed together with the Python implementation, which itself leans heavily on the use of NumPy concepts and machinery. A more language-agnostic approach to the core specification, together with some slimming down of the specification, would help to achieve complete implementations with full interoperobility across major programming languages. …*
- Vince has reached out to the community manager for Zarr for guidance on bindings that may already exist for v3.
- The scverse has a strong link to the Zarr group - connect via them?
- The first Visium HD example data set is available. Large parquet files - DuckDb or similar to work with these?

:24 - :25 Status update on the Bioc 3.19 (2024-05-01) release
- 3.19 released, a few post-release tasks in progress (e.g. Docker and webR rebuilding).

:25 - :27 Revisit idea of using CRANhaven (https://www.cranhaven.org/) to mitigate CRAN package dependencies that are temporarily archived on CRAN
- A note has been added in the BiocManager vignette.

:27 - :36 Suggestion: Don't bump version when there has been no updates (=no commits)
- The current software stack and practice of going from release to release are quite rigid (and work well).
- Long history and motivation behind the automatic bumps - odd/even version scheme, automatic branching. If automatic bumps and automatic branching are removed, many other things need to change. No room for bumping z in release if the devel y version has not been bumped.
- Even if the version stays the same, binaries may still be built for different versions of R.
- Automatic bumps introduce friction for the users. Don't know whether something was updated. NEWS files are often not up to date (especially with respect to automatic bumps).
- Could we add another element to the version (that would be appended by Bioc)?
  - What format could that be? "x.y.z+build" (e.g., 1.0.1+bioc318 vs 1.0.1+bioc319) is not valid. May have to be something like "1.0.1.319".

- Another reason for having the version bump is to allow identification of whether Bioc versions are mixed.
- How can we improve the way that changes and updates to packages are communicated?

:36 - :50 GitHub Actions and r-universe [document link, slides]
- Build system has been working in the current format for ~17 years. Are there alternative approaches?
- r-universe is handling ~20K packages.
- https://bioconductor.org/dashboard/
- Possibly separate a GitHub organization that does the builds (Enterprise), and another organization with the code (free).
- Want to provide actions to users - in smaller chunks.
- Differences in how logs are surfaced in r-universe and the current BBS reports.
- How to avoid propagating a version that is not working?
- The plan is not to take all of r-universe "as a package", but to adopt the aspects that are helpful, within the Bioc framework.
- Also probably don't want to propagate any change (e.g. without a version bump).

:50 - :55 Blog post plan
- Relation to Carpentries module, dialogue about the Carpentries module at TAB.
- Explain why we do things the way we do.

:55 - :60 AI and other research items
- E.g. https://github.com/biomap-research/scFoundation/tree/main
- Sean: We are starting to think on our campus about a broad LLM strategy. This is a bit dated, but it has some useful information. Would, of course, love contributions. https://seandavi.github.io/campus-llm-kb/ The research page is the most fleshed out: https://seandavi.github.io/campus-llm-kb/resources_education.html