Bioc Technical Advisory Board Minutes

7 September 2023

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

:00 - :18 Welcome to new members, information
- Should we do breakout rooms in future TAB meetings?
- Is anyone working on pangenomes/graph genomes? Should we be looking at new genome representations? Discussion with Heng Li at BioC2023 suggested that at least for a while we'll probably project back to a linear genome (T2T?), and may not need to overhaul e.g. how we do overlaps. Tim Triche is looking at pangenomes in cancer space, and may have more insights.
- NIH U24 funding opportunity for knowledge bases (documentation, workflows?) (https://grants.nih.gov/grants/guide/pa-files/PAR-23-237.html)
  - Mike is part of the IGVS consortium - building a knowledge graph in a graph relational database. How can we have programmatic access to that?
- Welcome to new members: Helena Crowell, Henrik Bengtsson, Ludwig Geistlinger.
- Brian Schilder is new chair of the Cloud Working Group.
  - Working group chairs are invited to report on group activities.
- TAB purpose (from website About page link): The Technical Advisory Board purpose is to support the Bioconductor mission by
  - Developing strategies to ensure long-term technical suitability of core infrastructure for the Bioconductor mission. Core infrastructure includes: all aspects of package addition, management, and distribution; end-user engagement (e.g., web, support site, and slack); developer support; and development of packages for use by the broader developer community.
  - Identifying and pursuing technical and scientific aspects of funding strategies for long-term viability of Bioconductor.
- Would like TAB to actively propose discussion topics in computation and genomic data science for each meeting, with topics for the next six months spelled out soon.
- New CZI EOSS cycle begins, LOI Oct 17.
  - S4Vectors/Delayed/Biostrings
  - Workshop platform
  - More support for the OSCA book (maintenance + new features, multimodal, cell type annotation). What kind of budget would be proposed?
  - Johannes and Laurent may put together a proposal about mass spec
  - Extensions to iSEE for spatial omics
  - Would we engage NumFocus?

- - Coordinate among Bioc-related proposals. The project could provide letters of support.
  - [Complex NSF opportunity announcement](#)

:18 - :23 Code of Conduct revision proposal
- "The Bioconductor community is committed to ethical partnerships and avoiding partnerships with entities whose practices are inconsistent with Bioconductor values. We encourage community members to report or raise concerns about partnerships they deem unethical to the CoC committee via the reporting form."
- An internal procedural guidance document section should be available for all Code of Conduct clauses. What is the review process for a complaint, what decisions are rendered, what are the options for responsive action?
- The CoC committee has an incident response guide.
- The TAB approves the formulation.

:23 - :38 Low-level technical concerns
- _R_CHECK_SUGGESTS_ONLY_ leads to failures for many packages. Should we enforce it?
- Approach to defunct (bouncing) emails for packages, where we don't know how to get in touch with the maintainer.
  - For example, aCGH has lots of downstream dependencies but emails bounce.
  - Have tried to reach out also to additional authors, if available.
  - Put a badge on the landing page: "Non-responsive maintainer". Automatic deprecation can have big downstream effects.
  - Crowdsource search for maintainers (Lori has a list).
  - If we don't hear from the original authors, an interested community member can take over maintenance.
  - Add regular automated checks of email addresses in DESCRIPTION files? Require the maintainer to confirm that the package is still being maintained.
- workshops.bioconductor.org
  - Need for backup. Azure, Orchestra are options. Automated failover (unlikely)? Committed authorized tech support? Easy switch, health check.
  - Basing on Azure as a platform is not feasible because of security concerns. Erdal has allowed a setup as an emergency backup.
  - Plan: keep the Azure server as a development/test/backup server. Need to know what expenses we could face (probably similar to the cost for Orchestra).
  - Building on AWS? Can we use usegalaxy.eu as backup?
  - Role of the containers that are used in the background (could be used locally)?
- Release schedule for Oct/Nov 2023 is available.
- Non-standard OrgDb (1500+ updated every release)
  - Uses AnnotationDbi functionality - also needs an overhaul, efficiency could be improved.
  - A lot of memory is required for the generation.
  - Non-model organisms, OrgDbs change regularly.

- - - Could they be generated on-demand?
    - What level of transformation is involved? Can we interface to the original data source?
    - This is a very valuable resource, removing them may have a large impact on people's work. The intention is not to remove them.
  - Hubs renovation waiting on AWS modernization (can't use OpenAPI with Ubuntu 14.x).

:38 - :41 Rsamtools
- Doesn't handle CRAM, some code approaching it is available.
- Are there important use cases for CRAM processing?
- PR from Brian Schilder exposes some C-level warnings, provides GitHub action.
- No named owner/maintainer ("maintainer@bioconductor.org"). Also the case for other core packages.
- Rhtslib updating process?
  - Some warnings were propagating.
  - htslib devs did make source changes - not enough to warrant a full update at Bioc for Rhtslib.
  - Training opportunity for interested individuals - if you or a colleague want to learn how to upgrade C-based components of an R package with potentially large user base, let us know.

:41 - :42 BiocBook
- https://jserizay.com/BiocBook/devel/
- Simplifies the production of a book.
- Laurent will test it with the R for Mass Spec book.
- There will be a BoF session at EuroBioC.

:42 - :48 r-universe
- https://r-universe.dev/search/
- Contains all Bioc packages available on GitHub? https://r-universe.dev/search/?q=bioconductor
- Release branch? https://bioconductor.r-universe.dev/BiocParallel
- 45 Bioc packages presented as organization: https://bioconductor.r-universe.dev/builds
- Are binaries compatible e.g. with Bioc containers? https://github.com/r-universe-org/help#does-r-universe-have-linux-binaries
- Determination of upstream source: https://ropensci.org/blog/2023/04/03/cran-to-git/
- Can we make use of any of the underlying technology (CI/CD)?
- Are they running tests?

:48 - :49 Hacktoberfest
- https://hacktoberfest.com/
- October 1 - October 31
- Annual event that encourages people to contribute to open source.
- To participate as a maintainer:

- - ○ Prepare repositories by adding the 'hacktoberfest' label.
    - ○ Prepare suitable issues.
  - ● Possible to organize events (perhaps related to a specific topic/repo/theme).
  - ● #hacktoberfest slack channel.

:49 - :60 Technical opportunities
- ● ArtifactDb in ITCR U24
  - ○ Language agnostic annotation etc. – what is a roadmap for credentialed/automated contributions?
  - ○ alabaster is already there.
  - ○ CollaboratorDb has been made available but we need to learn more about it.
- ● SOMA TileDb for HCA, other HCA interop.
- ● SparseArray progress - abind working well, Rafa's use case? - 90% feature complete, need more use cases.
- ● ML/AI/LLM opportunities
  - ○ keras model handling, orthos.
  - ○ BiocSfaira.
  - ○ [Sean's gist](#) to use Chat-GPT to create a codebook for a track from AnnotationHub, augment and standardize metadata.
  - ○ Assistance in writing R code and documentation.
  - ○ Alex is working on a (python) package that would do test-driven development (turn test cases into code).
- ● New website and related opportunities for identity management and customized user experience.
  - ○ Deployment is coming up in the near future.
  - ○ The plan is to keep the old site and new one in parallel to clean up content issues before having the new page as the main one.
- ● Communication channel consolidation
  - ○ Bring the bioc-devel mailing list and slack together? Automatic forwarding of bioc-devel emails to a slack channel. Not everyone is on slack - how to make sure information gets to everyone (if it's only replied to on slack)?
  - ○ slack is more searchable.
- ● Automated summaries of meetings? Testing it for this meeting.
- ● We received a grant supplement - may allow purchase of MacPro for builder, Linux machine with GPU?
- ● ARM Linux builds – "experimental" but could last longer.
  - ○ Should the badge be sensitive to this?
  - ○ We will provide binaries for this platform - via GHA emulator