Bioc Technical Advisory Board Minutes

1 October 2020

Attending: Stephanie Hicks (joined at :30), Vince Carey, Laurent Gatto, Martin Morgan, Sean Davis (guest), Charlotte Soneson, Robert Gentleman, Shila Ghazanfar, Wolfgang Huber, Michael Love, Levi Waldron, Aedin Culhane, Hector Corrada Bravo
Regrets: Aaron Lun, Kasper Hansen, Rafael Irizarry

:03-:05 - 2020-09-03 minutes approved; governance process progress

:05-:15 - Greatest hits of September 2020 (Suggestion: Can we envision an approach to tracking the project at this granularity, in a coordinated way?)
- Windows builds: workflows and experiment packages running again after long hiatus
- CRAM ingestion for Rsamtools (work in progress)
- SparseMatrixStats discussion -- seems ripe for a demonstrative workshop/developer forum -- absolutely needed this year. Recommendation to invite Martin Maechler and Henrik Bengtsson.
- Bioconductor on Azure -- thanks Nitesh
- Conrad Burden: technology overload for simple changes to fix a package -- bioc-devel mailing list.
    - could/should we have a web interface to the git server? Possible issues with having interaction (issue tracking etc) in multiple places.
- velociraptor and Github Actions resource limits
- biocthis -- Leo Collado Torres has been invited to give the tech talk for November TAB meeting
- September developer forum rapidly organized to present AnVIL package and containerization/repository concepts, with extended discussion by Dirk Eddelbuettel; container maintenance discussion related to current OpenSSL
- Progress with k8sredis -- Bioc infrastructure driving and profiting from Kubernetes
- How should we deal with conference workshop materials? They will work with the paired container but can still go stale. After 3.12 is released, should workshop Rmd be tested against it?
    - with GitHub Actions, maintenance and development is in the hands of the developer. Can generate new docker images with a scheduled GitHub Actions run; the workshop platform will use the updated docker image.
- SAB recommendation: package review process should be driven by community and core developers can reduce engagement with it -- could we implement this?
    - try a pilot phase first (a few packages, contributors opt-in?)
    - deserves a working group with membership from CAB and TAB to hash out and oversee implementation

- - develop a rubric (see e.g. [rOpenSci](#) material) with the aim of standardization across reviews - also for evaluating the review itself
    - continuous model of review rather than the current "gatekeeper" review model? How to keep motivation up?
    - different options possible: editor/associate editor approach, or more of a community approach.
- Ludwig Geistlinger [query](#) on multiple experiment datasets -- ExperimentHub a good enough answer?
- Ongoing discussion of application to NSF Cyberinfrastructure for Sustained Scientific Innovation framework [solicitation](#)
- Annotation (from Aedin) - With more single-cell data, or sequence data that has genome coordinates, we are no longer supporting a myriad of oligo. Given that annotation is theoretically simpler, could we again consider auto filling or standardize min format for rowData (EnsEMBL, Symbol, Chr start/end/strand etc)?

:15-:30 - CAB liaison
- Community commitment to package review process, to free up core developer time - build up an explicit rubric for reviews (cf. [rOpenSci](#)). This could perhaps open up other possibilities: for example, packages accepted after rOpenSci review are [fast-tracked](#) at [JOSS](#).
- CAB Minutes from [8/13](#).
- Topics under discussion:
    - governance discussions
    - racial diversity statement
    - funding proposals/budget
    - possibilities to standardize software/tools used for meetings?
    - [EuroBioc](#): will use hopin.to (closed captioning not supported, only Google Meet does this so far as we know). OpenReview.net ([link](#)) used for submissions/reviews. Short days (afternoons). Can we use the meeting to catalyze development on some of the greatest hits above ([submit proposal](#)!)?

:30-:50 - Sean Davis presentation on Orchestra, the infrastructure used for the BioC2020 workshops
- Slides: https://www.dropbox.com/s/a5vwbjfk4fca4pd/2020-10-01-orchestra-bioc-tab.pptx?dl=0
- Software: https://github.com/seandavi/Orchestra
- Discussion:
    - 'Service' or 'software'? Rather a service - the user creates a Docker image, Sean adds it to the collection.
    - Is the need to create a Docker image a burden for the workshop creator (it's unrelated to the topic of the workflow)? Seemed to work well for most BioC2020 workshops, difficult to find a good alternative. For many Bioc workshops it might

be enough to use the basic bioconductor docker image, and let the participants install the required packages.
- Could potentially provide a 'build system' inside Orchestra in the future
- Docker images for Bioc workflows? Works fine as long as they are set up as R packages.

:50-:60 - Open floor, please add topics.
- How can we establish topics (and larger audience) for developer forum?
  - slack poll
  - make it easier for people to volunteer. Rotating moderators?
  - is the name "developer forum" intimidating? Rename to "Developer and community interest forum" or "Contributor forum"?
  - small awards to engage people?
- There was no closure on September discussion of strategies for "book" production and maintenance; OSCA is clearly a dynamic resource, with new information on handling [ambient RNA](#) forthcoming. Need for unit tests or regular builds to alert of content updates (Amazon Lambda services?)
- Similarly for writing groups for topics like *hub, BiocParallel/GenomicFiles, workshop platform
- https://aws.amazon.com/quickstart/biotech-blueprint/ Call today (2020-10-01) at 4pm EST. Just introductions.

Appendix: Thoughts on a working paper concerning the build system for our multifaceted ecosystem -- something VC has to do for CZI BBS EOSS grant. Contributors will be sought out but any comments/co-authors welcome
- Introduction: scope of the project
  - open
  - portable: to 3 platforms
  - agile: track R, biotech
  - reach research, education, publishing
- User views:
  - starting and extending the workspace: BiocManager
  - long term projects
- Developer views
  - using components of the ecosystem
  - adding a package to the ecosystem -- criteria
  - managing a contributed package
- Technical underpinnings
  - The released ensemble of packages and their non-Bioc dependencies
  - enumeration
  - configuration of a linux builder
  - population and maintenance of a repository
- System health reporting

- ○ R CMD check
  - ○ R CMD BiocCheck
  - ○ test coverage
  - ○ build reports
  - ○ emulating the private build system by independent developers
- Evolving the ecosystem: package removal, API deprecation, package review and admission
- Paths forward
  - ○ Governance and distribution of tasks
  - ○ Containerization, binaries
  - ○ Increasing scalability: k8sredis
  - ○ Agility for developers: biocthis
  - ○ Sub-ecosystems: books and workshops