# Bioc Technical Advisory Board Minutes

5 November 2020

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

:03-:05 2020-10-01 minutes approved

:05-:07 A systematic approach to project strategy, priority adoption, and conflict resolution is needed to ensure sustainability of the organization.
- Example 1: the project needs to consider how to manage/sustain growth in scope of software package coverage, innovation in delivery of exemplary data and annotation, and service to developer and user communities. The need/opportunity for multiple funding streams (government research grants, other foundation grants, industry contributions) is not easy for full-time investigators to manage.
- Example 2: adoption of project-wide issue tracking.
- Example 3: feature planning for key infrastructure. This diagram shows some of the issues at stake as we must be concerned about overreach.

Comments: Fundamentally a matter of resources. Code moratorium to discuss priorities? Incentive structure. How can TAB help incentivize contributions?
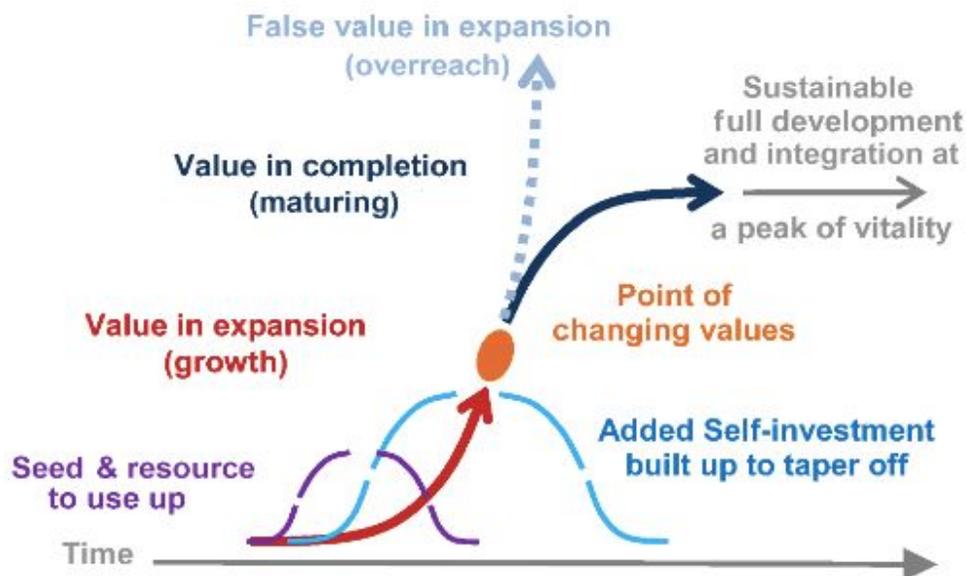
Source: Tainter

:07-:10 Governance. CHAOSS (EOSS grantee) could be useful as a source of comparable projects

:10-:20 Greatest hits of October 2020

- Freeze of 3.11/release of 3.12 . Major activities [Thanks to Martin Morgan for summary]:
  - A push to review submitted packages in time for the release -- 125 (a record) packages added this release.
  - Divide packages into 'ranges' assigned to each core member; core members responsible for reaching out to packages failing across platforms to engage developers in fixing the problem. Complements ongoing quasi-automatic identification of failing packages as candidates for deprecation. Deprecating and removing no-longer-functional packages is increasingly important.
  - Rebuild 'annotation' resources. While the core team has enough expertise to 'turn the crank', we employed James MacDonald to provide a more robust update to the workflow. James has identified several significant areas (e.g., we produce out-of-date UniProt data resources) that could form the basis of additional expert work.
  - Reconfigure build machines to support next 'devel' builds, including updating OS (to Ubuntu 20.04, for instance) as well as ensuring sufficient disk space / compute power to handle ever-increasing demands.
- Support site renovation
  - Huge thanks to Natay Aberra, working with Istvan Albert, for both the update and very conscientious response to pre- and post-release bug reports
- Project-wide code of conduct
  - Interest in [translation](#) to other languages
  - CoC enforcement concepts are also not clear. Will the CoC committee be fully empowered to carry out enforcement? How to nip conflicts or problematic threads in the bud, and get buy-in from all community members that this approach to problem management is acceptable?
  - Slack private channels/messages
- Johannes Rainer's https://jorainer.github.io/SpectraTutorials/ with docker (suggestion: TAB tech talk on proteomics infrastructure/analysis tools)
- Mike Lawrence's [answer](#) to Laurent on List():
  - There's an opportunity to implement faster matching than base::merge(), using stuff like matchIntegerQuads(), findMatches(), and grouping().
  - grouping() can be really fast for character vectors, since it takes advantage of string internalization. For example, let's say you're merging on three character vector keys. Concatenate the keys of 'y' onto they keys of 'x'. Then call grouping(k1, k2, k3) and you effectively have a matching. Should be way faster than the paste() approach used by base::merge(). Would be interesting to see.
- Constantin Ahlmann-Eltze's [developer forum](#) on sparse matrix stats
- Change management -- how should we deal with workshop materials? They will all work with the assigned container but they can still go stale. Proposal: after 3.12 becomes release, the workshop materials should test against it.

- https://kevinrue.github.io/BiocChallenges/ - collection of 'challenges' related to Bioconductor, **compiled for EuroBioc2020**. Could turn into a useful platform for collaborative work.
    - TAB could revisit on a regular basis
    - Scope and engagement of general core activities
    - Each challenge has a leader identified who can answer questions but need not do the technical work on the solution
- Hervé [started to work](#) on incremental builds every 4 hours: "Every 4 hours we'll build and check packages that have changed since the last run. Will be on nebbiolo1 (Linux) and riesling1 (Windows) only for now. The other builders don't have enough power to handle the extra load. These builds will be informative only. No binary builds, no propagation."
- New slack channels #spatialexperiment and #microbiomeexperiment for discussing development of containers for spatial transcriptomics and microbiome data.
- Book production has gone [live](#)/mainstream
- Pain points:
    - dropbox/bitbucket/... endpoints for downloading resources -- does BiocCheck flag? … needs strategic alignment of core and developer interest -- Rirods, globus -- transferring large data
    - Aaron Lun -- "[EHub side issue (Ahub)?](#)" bfcadd failing related to AWS availability

:20-:30 CAB liaison
- Community commitment to package review process, to free up core developer time … build up a rubric for reviews that is as explicit as the rOpenSci. Work in progress.
- [CAB September Minutes](#)
- TAB Stickers are available.
- F1000Research re: ideas for relaunch of BioC Gateway. Bioconductor project turns 20(?) in 2021.
- List of seminar series presenters for H3ABioNet (available [here](#))
    - Yagoub will finalize it by Monday 12/10/2020
- [BioC Asia 2020](#) (October 15-18 (half days), virtual format) 441 people registered
- [BioC Europe 2020](#) (December 14-18, virtual format)
- [Bio"Pack"athon](#) ([October 14 [Japan time], virtual format](#))
- Geographic slack channels including [#bioc_africa](#), #biocasia. Do we need a guide to Slack Channels?

:30-:60 Leo Collado Torres on biocthis [[slides](#)]
- Fill 'holes' between what you get from usethis and a package that you can submit to Bioconductor. usethis-like functions that fulfills Bioconductor's needs (including vignette)
- Leverage/set up GitHub Actions workflow (including Bioc docker containers + macOS + windows) - Bioconductor-friendly setup (R CMD build + check + BiocCheck)
- On GHA (unlike BioC build system), increased exposure to changes in R build issues, dependency management - complementary, not a replacement for the nightly Bioc

builds. Useful for increasing productivity, noticing issues faster, helping new users become developers.

- May be an important building block if/when moving to a containerized build system.
- Maintenance problem of the generated YAML? Recommended solution is to rerun biocthis regularly to overwrite the generated YAML file.