Bioc Technical Advisory Board Minutes

6 March 2025

Present: Vince Carey, Lori Kern, Davide Risso, Erdal Cosgun, Jacques Serizay, Charlotte Soneson, Laurent Gatto, Helena Crowell, Alexandru Mahmoud, Robert Gentleman, Sehyun Oh, Marcel Ramos, Hervé Pagès, Ludwig Geistlinger, Michael Lawrence, Andres Wokaty, Levi Waldron, Rafael Irizarry, Henrik Bengtsson, Stephanie Hicks
Apologies: Wolfgang Huber, Kasper Hansen

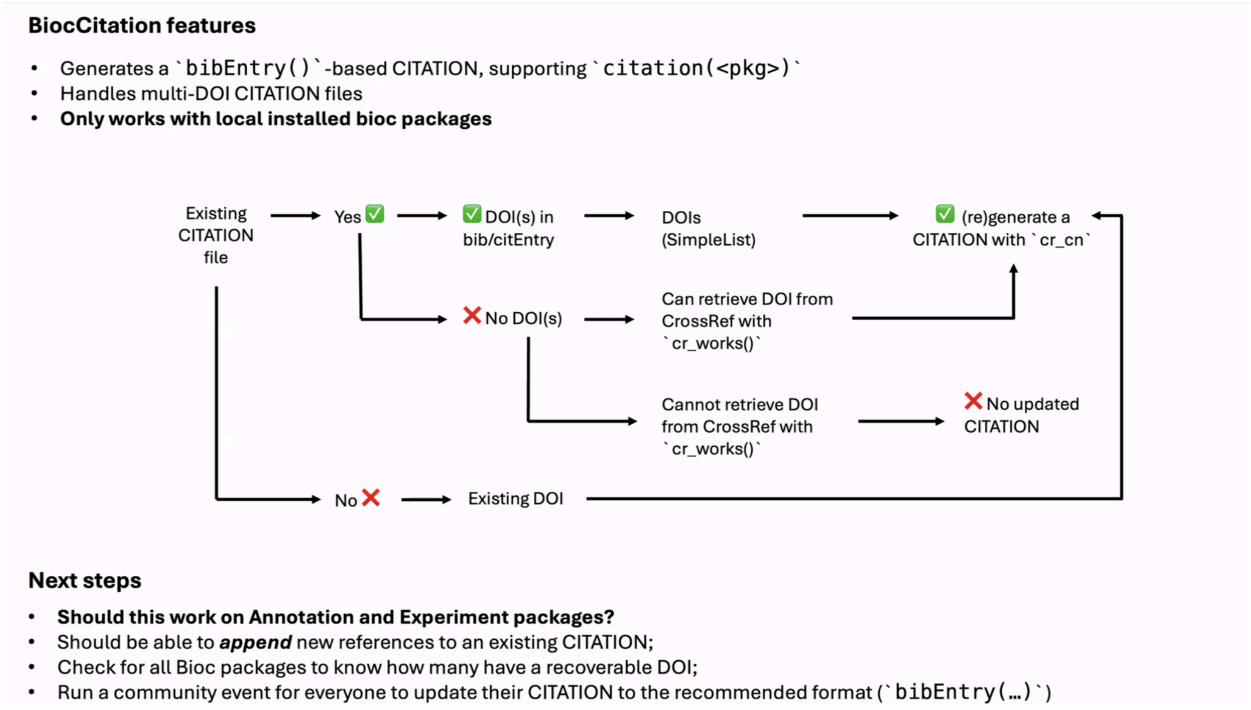
:00 - :05 Previous minutes approved; overview of working groups.

:05 - :14 Discussion about the current state of the NIH and associated uncertainties.

:14 - :17 CAB update
- New CAB meeting time approved
    - 2 pm GMT Apr - Oct
    - 10 pm GMT Nov - Mar
- Joint CAB/TAB meeting next month during regular CAB meeting
    - April 10th 2025, 2 pm GMT (9 am ET).
    - Lori will send an agenda out shortly for people to add topics or reach out to Lori via email or slack to add items.
- Meeting recap
    - More members introduced what initiatives they would like to be a part of.
    - BiocClasses discussion.
    - CAB Governance discussion and voting on updates.
    - BiocAwards are paused for this year because of joint conference with Galaxy.
    - CAB Elections will not coincide with conference due to 2 month move up - will still hold elections in August.

:17 - :20 Upcoming release
- https://bioconductor.org/developers/release-schedule/
- April 16th target release date for Bioc 3.21.
- Annotation pipeline warrants significant effort - human.db0, other db0, org…, TxDb….
    - Manufactured using 15+ years old code, very fragile elements are present, dependent on external services consistency and stability … could the fragmentation across organisms/data types be reduced?

:20- :25 CITATION infrastructure (Jacques)
- BiocCitation package initiated.
- Generates a bibEntry()-based CITATION.



BiocCitation features

- Generates a `bibEntry()`-based CITATION, supporting `citation(<pkg>)`
- Handles multi-DOI CITATION files
- **Only works with local installed bioc packages**

Existing CITATION file
Yes ✅
✅ DOI(s) in bib/citEntry
DOIs (SimpleList)
✅ (re)generate a CITATION with `cr_cn`
❌ No DOI(s)
Can retrieve DOI from CrossRef with `cr_works()`
Cannot retrieve DOI from CrossRef with `cr_works()`
❌ No updated CITATION
No ❌
Existing DOI

Next steps

- **Should this work on Annotation and Experiment packages?**
- Should be able to ***append*** new references to an existing CITATION;
- Check for all Bioc packages to know how many have a recoverable DOI;
- Run a community event for everyone to update their CITATION to the recommended format (`bibEntry(…)`)

:25 - :33 Image analysis working group (Sehyun)

- There is a minimal connection between image analysis software (mostly Python) and downstream statistical analysis software (mostly R/Bioconductor).
- Aims: Integrate state-or-the-art non-R image analysis software with the R/Bioconductor ecosystem (develop a pipeline from raw histopathology image to image-multiomic integration in R/Bioconductor), create a repository of histopathology image features from TCGA image.
- Goals: develop data structures and establish standardized workflows for importing image features from Python-based tools into R/Bioc data structures, process TCGA histopathology images and create a curated repository, develop training material and documentation.
- Current pipeline starts from image slide, preprocess using PathML, Hover-Net to extract features (coordinates for nucleus boundary, centroid, contour) -> .json file -> AnnData -> extract more features (shape, texture) using squidpy, scikit-image. Plan to add extraction of ROI-level features.
- `imageTCGA` package: shiny app for exploring TCGA diagnostic image database.
- Hold biweekly meetings.

:33 - :46 Other working groups

- Classes and Methods: looking into potential for improving documentation, guidance for submissions, recommendations around S7, feasibility to expand S4/S7 compatibility, S4Examples and S7Examples packages under development.
- Training: https://github.com/Bioconductor/BiocHowTo (contributions welcome), training paper was accepted for publication.

- Mass Spectrometry: efforts to integrate R and Python infrastructure for Mass Spec ongoing (https://github.com/rformassspectrometry/SpectriPy).
- Spatial: working on OSTA book (with a group partly overlapping with this working group).
- Multilingual: working session upcoming to translate some lessons to French. EU funding ("science in your own language" - https://www.chistera.eu/) available - discussions about possible projects ongoing.
- Workflows: had a call with PCI (peer community in), exchange with JOSE - aim to find new outlets for workflow publishing.

:46 - :60 Other topics

- Perils of slots for analysis outputs like reducedDims (how does subsetting with [,] avoid confusion?), example. More generally: how to be sure of provenance, is there a contract?
- Is there a role for the targets package in organizing/caching multistep/multisubset activities in a more resilient manner?
- https://github.com/jackgisby/ReducedExperiment
- Track/provide the code used to derive the reduced dimensions.
- Could drop derived slots (but this would break a lot of code).
- Not just an issue with reduced dimensions - also row/colData columns calculated from a larger/smaller object.