Common Repository Working Group

2019/2020 Update, 2021 Plans

AIRR Common Repository Working Group (CRWG)

- Purpose
  - ***To promote and facilitate deposit, access, and sharing/reuse of IG and TCR AIRR-seq datasets*** through the creation of common repositories that enable:
    - standardized queries of processed AIRR-seq data.
    - re-analysis of raw and processed AIRR-seq data utilizing repository analysis tools.
    - download of raw and processed AIRR-seq data for offline re-analysis.
- Who we are:
  - Co-chairs: Brian Corrie, Lindsay Cowell
  - Members: Meredith Ashby, George Blanck, Felix Breden, Tania Bubela, Syed Ahmad Chan Bukhari, Christian Busse, Brian Corrie (Co-lead), Scott Christley, Lindsay Cowell (Co-lead), John Harting, David Klatzmann, Uri Laserson, Bjoern Peters, Artur Rocha, Adrien Six, Adrian Thorogood, Eric Waltari, Corey Watson, and Yariv Wine.

Plans for 2019/2020 (from Genoa meeting)

- Recommendations: Continue to adapt as required
- Finish AIRR API specification
- Ontologies: Continuing to work on Ontology sub-working groups
- Registry: Programmatic registry that is searchable
- Meta-analysis working group
- Manuscript - AIRR Community endorsed

AIRR CRWG Recommendations - Updates

- AIRR 2020 Meeting CRWG Recommendations
  - https://github.com/airr-community/common-repo-wg/blob/airr-2020/recommendations.md
- Housekeeping
  - Shortening of recommendation titles
  - Consistent use of terminology (e.g. "repository" over "database")
  - Addition of explicit mention of standards (MiAIRR, DataRep, ADC API)
- Structural changes
  - R5 & R6 - Order reversed, minor wording changes
- Updates to "Next Steps" match 2021 plans
- ***Future: Continual updates as required***

AIRR Data Commons (ADC) API

- Web API: Programmatic mechanism to query/download AIRR-seq data
- ADC API v1.0 released
- ADC API Manuscript (Frontiers in Big Data - DOI: 10.3389/fdata.2020.00022)
- ***Future: Extending ADC API***

TECHNOLOGY AND CODE ARTICLE
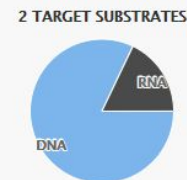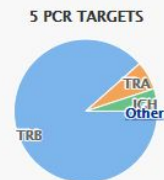
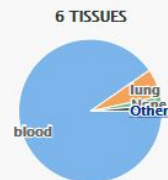Front. Big Data, 17 June 2020 | https://doi.org/10.3389/fdata.2020.00022

The ADC API: A Web API for the Programmatic Query of the AIRR Data Commons

Scott Christley[1], Ademar Aguiar[2,3], George Blanck[4], Felix Breden[5], Syed Ahmad Chan Bukhari[6], Christian E. Busse[7], Jerome Jaglale[5], Srilakshmy L. Harikrishnan[7], Uri Laserson[8], Bjoern Peters[9,10], Artur Rocha[2], Chaim A. Schramm[11], Sarah Taylor[12], Jason Anthony Vander Heiden[13], Bojan Zimonja[5], Corey T. Watson[14], Brian Corrie[5] and Lindsay G. Cowell[1*] on behalf of The AIRR Community[†]

AIRR Data Commons

- AIRR Data Commons
  - Network of AIRR Compliant Repositories (AIRR Standards, ADC API)
  - Three large repositories (AIRR COVID-19, iReceptor Public Archive, VDJServer)
    - 4 billion annotated sequences, 6013 repertoires, 60 studies, 3 repositories
  - Significant COVID-19 data
    - 1 billion annotated sequences, 3489 repertoires, 19 studies, 2 repositories

- *Future: Advancing AIRR Data Commons*
  - *More data, new repositories, community engagement*

1,070,305,731 **sequences (3489 repertoires)** returned from 2 remote repositories, 13 research labs and 19 studies.

4 STUDY TYPES	2 ORGANISMS	2 DIAGNOSES	6 TISSUES	5 PCR TARGETS	2 TARGET SUBSTRATES

Observational Study · Case-Control · Longitudinal Study · Study

Homo sapiens · Mus musculus

COVID-19 · None

blood · lung · None · Other

TRB · TRA · IGH · Other

DNA · RNA
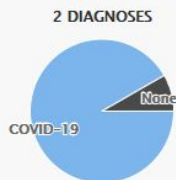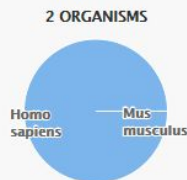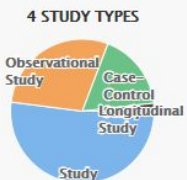
Ontologies

- Sub-working group struck to focus on Ontology use
  - Spanned many working groups, sprint focused (short "vigorous" work periods)
- Sprint - 04/2020
  - See full Sprint Report on AIRR Community Docs and MinStd report
- ***Future: Continuing to work on Ontology sub-working groups***

AIRR-seq Repository Registry (AIRRrr) ☺

- It is hard to find repositories that store AIRR-seq data!
- What is an AIRR-seq repository registry?
  - A place to **discover** AIRR-seq data repositories
- Current registry (manual discovery!!!)
  - AIRR Data Commons documentation page
  - AIRR-seq data repositories on b-t.cr (not all AIRR compliant)
- 2019/2020 plan: Programmatic registry that is searchable
  - Did not progress on this task in 2019/2020
- *Future: Programmatic AIRR Repository Registry*

AIRRrr!
There be
Treasure!

Meta-analysis working group

- Proposed/discussed at Genoa meeting
  - Not enough interest, so did not proceed...

Leadership - co-chairs

- Leadership changes
  - Remaining: Brian Corrie, SFU
  - Leaving: Lindsay Cowell, UTSW (AIRR Exec Chair)
  - New Co-chair: Artur Rocha, INESC TEC, Portugal

AIRRrrr!
There be treasure...

Big Picture - AIRR CRWG Plans 2020/2021

- Outward focus - Focus on promoting/advancing the AIRR Data Commons

- Moving standards/specifications work to new Standards Working Group

- Please join us!!!

Summary - AIRR CRWG Plans 2020/2021

- Outreach to community to expand the AIRR Data Commons
- Manuscript - AIRR Data Commons from the user perspective
- Outreach to Journals and Funders to adopt AIRR Standards/Data Commons
- Recommendations: Continue to adapt as required
- Registry: Programmatic registry that is searchable
- Standardized consent for data sharing (with Legal & Ethics WG)
- Standards WG related plans
    - Continue to evolve the ADC API specification as the AIRR specification evolves
    - Ontologies: Continuing to work on Ontology sub-working groups
    - Persistent identifiers: PIDs required for the AIRR Data Commons