

AIRR Data Representation Working Group

2020 Update, AIRR-C V

Jason Vander Heiden & Scott Christley (*co-chairs*)

Purpose

Define data representations that make it easy to analyze, share, and inter-operate with AIRR-seq data.

Goals for 2019–2020

- 1) Community outreach and schema adoption
- 2) Refine existing standards, tools, and documentation
- 3) Enable other working groups
- 4) Develop new standards

1) Community outreach - Rearrangement

2018: 11 tools supporting, 4 planned

2020: 13 new tools supporting

TABLE 1 | Tools and databases supporting the AIRR Rearrangement schema.

Software	Version	Support
AIRR Python Library	1.2	Input, output and validation
AIRR R Library	1.2	Input, output and validation
IgBLAST	1.10	Output
IGoR	TBD	Input and output
Immcountant:Change-O	0.4.2	Input, output and conversion
ImmuneDB	0.24.0	Output
iReceptor	2.0	Input, output and conversion
MiXCR	2.2.1	Output
OLGA	TBD	Input and output
Partis	TBD	Output
SONAR	3.0	Output
TRIgS	2	Input
VDJServer	1.2.0	Input and output
Vidjil-algo	2018.10	Output
Vidjil Web Platform	TBD	Input and conversion

Software	Version	Support
Alakazam	1.0.1	Input and output
Cell Ranger	4	Output
Decombinator	4.0.1	Output
IMG/HighV-QUEST	1.7.0	Output
IMG/V-QUEST	3.5.16	Output
immunarch	0.6.5	Input
immuneSIM	0.8.7	Output
RAbHIT	0.1.5	Input and output
scirpy	0.3	Input
SCOPer	1.0.1	Input and output
SHazaM	1.0.0	Input and output
sumrep	1	Input
TlgGER	1.0.0	Input and output

2) Refine documentation

Redesigned and significantly expanded the AIRR Standards documentation site:

<https://docs.airr-community.org>

- Improved readability, accessibility, and navigation.
- Improved automation of builds and tests.
- Added detailed release notes.
- Encompasses documentation for multiple working group products, including:
 - MiAIRR Standard
 - Public Repository Data Submission Guides
 - Data Representation Schemas
 - AIRR Data Commons API
 - Software Guidelines
 - Ontologies & Vocabularies
 - Resources and Tools Supporting AIRR Standards



AIRR Standards 1.3
documentation

Q. Search the docs ...

Getting Started

Release Notes

AIRR Standards

Data Submission and Query

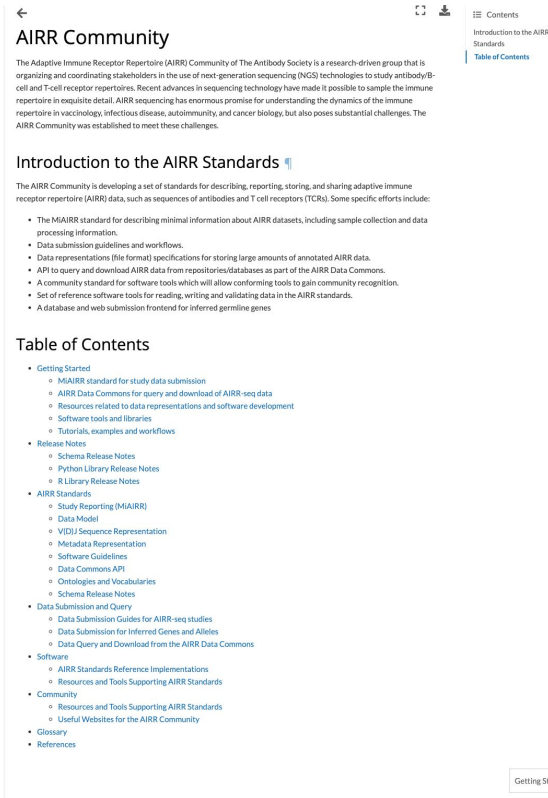
Software

Community

Glossary

References

Theme by the Executable Book Project



← AIRR Community

The Adaptive Immune Receptor Repertoire (AIRR) Community of The Antibody Society is a research-driven group that is organizing and coordinating stakeholders in the use of next-generation sequencing (NGS) technologies to study antibody/B-cell and T-cell receptor repertoires. Recent advances in sequencing technology have made it possible to sample the immune repertoire in exquisite detail. AIRR sequencing has enormous promise for understanding the dynamics of the immune repertoire in vaccinology, infectious disease, autoimmunity, and cancer biology, but also poses substantial challenges. The AIRR Community was established to meet these challenges.

Introduction to the AIRR Standards

The AIRR Community is developing a set of standards for describing, reporting, storing, and sharing adaptive immune receptor repertoire (AIRR) data, such as sequences of antibodies and T cell receptors (TCRs). Some specific efforts include:

- The MIAIRR standard for describing minimal information about AIRR datasets, including sample collection and data processing information.
- Data submission guidelines and workflows.
- Data representations (file format) specifications for storing large amounts of annotated AIRR data.
- API to query and download AIRR data from repositories/databases as part of the AIRR Data Commons.
- A community standard for software tools which will allow conforming tools to gain community recognition.
- Set of reference software tools for reading, writing and validating data in the AIRR standards.
- A database and web submission frontend for inferred germline genes

Table of Contents

- Getting Started
 - MIAIRR standard for study data submission
 - AIRR Data Commons for query and download of AIRR-seq data
 - Resources related to data representations and software development
 - Software tools and libraries
 - Tutorials, examples and workflows
- Release Notes
 - Schema Release Notes
 - Python Library Release Notes
 - R Library Release Notes
- AIRR Standards
 - Study Reporting (MIAIRR)
 - Data Model
 - VDJ Sequence Representation
 - Metadata Representation
 - Software Guidelines
 - Data Commons API
 - Ontologies and Vocabularies
 - Schema Release Notes
- Data Submission and Query
 - Data Submission Guides for AIRR-seq studies
 - Data Submission for Inferred Genes and Alleles
 - Data Query and Download from the AIRR Data Commons
- Software
 - AIRR Standards Reference Implementations
 - Resources and Tools Supporting AIRR Standards
- Community
 - Resources and Tools Supporting AIRR Standards
 - Useful Websites for the AIRR Community
- Glossary
- References

Getting Started >>

2) Refine existing standards

AIRR Schema v1.3

 schristley released this on Jun 1 · [37 commits](#) to master since this release

Version 1.3 of the AIRR Schema

Major changes:

- First official release of the Repertoire metadata schema.
- First official release of the AIRR Data Commons API specification.
- Expansion of schema field annotations.
- Extensions to the Rearrangement schema for multiple D regions.
- New experimental schema for clones, lineage trees, and cells.

Full Release Notes:

- [Schema release notes](#)
- [Python library release notes](#)
- [R library release notes](#)

- Major released v1.3 of the AIRR Standards.
- Improved release process.
- Detailed patch notes for schemas, python library, and R library.

3) Enable other working groups

Integrated content of Common Repository, Software, and Germline Database WGs into the AIRR Standards documentation site:

- Data Commons API
- Software Certification Guide
- OGRDB Submission Guide

Worked closely with Minimal Standards on additional standards and schema development.

AIRR Standards

Information about all of the AIRR Community standards.

Table of Contents

- [Study Reporting \(MiAIRR\)](#)
- [Data Model](#)
- [V\(D\)J Sequence Representation](#)
- [Metadata Representation](#)
- [Software Guidelines](#)
- [Data Commons API](#)
- [Ontologies and Vocabularies](#)
- [Schema Release Notes](#)

4) Develop new standards

Draft lineage and single-cell schemas released in v1.3:

- Clone, Tree, and Node objects defining clonal clustering, lineage trees, and individual tree nodes, respectively.
- Cell and Receptor objects defining individual cells and receptors, respectively.

2021: Merge with MiniStd & ComRepo?

- MiniStd will **not** continue as a separate WG
- Instead, a **new** Standards WG will be proposed, which will absorb MiniStd, DataRep and some parts from ComRepo
- Why?
 - The work in these WGs will continue for at least another year
 - The products of these WGs are closely connected and overlapping
 - The same people, talking about the same things, in three different calls? Let's be more efficient! (We tried it for 6 months, it works.)
- Standards WG will take care of the AIRR data schema, common exchange format, ADC API, minimal reporting requirements, ontologies & vocabularies and persistent identifiers.
- ComRepo will continue, focusing on building the ADC (as a real thing, not just a protocol stack).

2021: Same as the old boss

- 1) Community outreach and schema adoption
- 2) Refinement of existing standards, tools, documentation
 - Official releases of lineage and single-cell schemas.
- 3) Enabling other working groups
- 4) New standards development
 - Manifest schema for packaging files together.
 - Schema for grouping Repertoire objects.
 - Redesign of data and sample processing schemas.

Please join us!

airr-standards@googlegroups.com

<https://docs.airr-community.org>

<https://github.com/airr-community/airr-standards>