A Metadata Model and Validation Framework for Immunological Data within NFDI4Immuno: Building a Foundation for Comprehensive Data Integration, Analysis, and Open Science

Vera Bockhorn[1], Sebastian Ferrara[1], Axel Ronald Schulz[1], Christian Busse[2], Hyun-Dong Chang[1,3]

1 German Rhematology Research Center, A Leibniz Institute, Berlin
2 German Cancer Research Center, In the Helmholtz Association, Heidelberg
3 Technische Universität, Berlin

The rapid advancement of single-cell technologies—including flow cytometry, mass cytometry, and single-cell sequencing—has enabled the generation of increasingly information-rich datasets, intensifying the need for robust data sharing and reuse within immunological research.

The National Research Data Infrastructure for Immunology (NFDI4Immuno) addresses this need by integrating immunological data and metadata. By harmonizing metadata standards, ontologies, data models, and programmatic interfaces in collaboration with other NFDI consortia, NFDI4Immuno promotes seamless discovery and cross-referencing of data. Central to the initiative is the implementation of data stewardship practices aligned with the FAIR principles and the support of user communities in leveraging these resources effectively.

The project establishes a federated network of institutional repositories and accompanying software pipelines for standardized deposition, annotation, and retrieval of immunological datasets, while ensuring proper access control and provenance tracking. This addresses current barriers caused by inconsistent sharing and annotation practices, mitigating data redundancy and enabling more comprehensive analysis across studies.

Here, we present the current state of our metadata model, defining a comprehensive metadata structure compliant with the General Data Protection Regulation (GDPR) that includes controlled vocabularies to enforce standardized data annotation. Our validation and submission process entails comprehensive quality checks of submitted metadata, meaningful error reporting, and efficient data file upload.

Ultimately, NFDI4Immuno seeks to foster a culture of Open Science and reproducibility, empowering the immunological community to derive deeper insights and accelerate research progress.