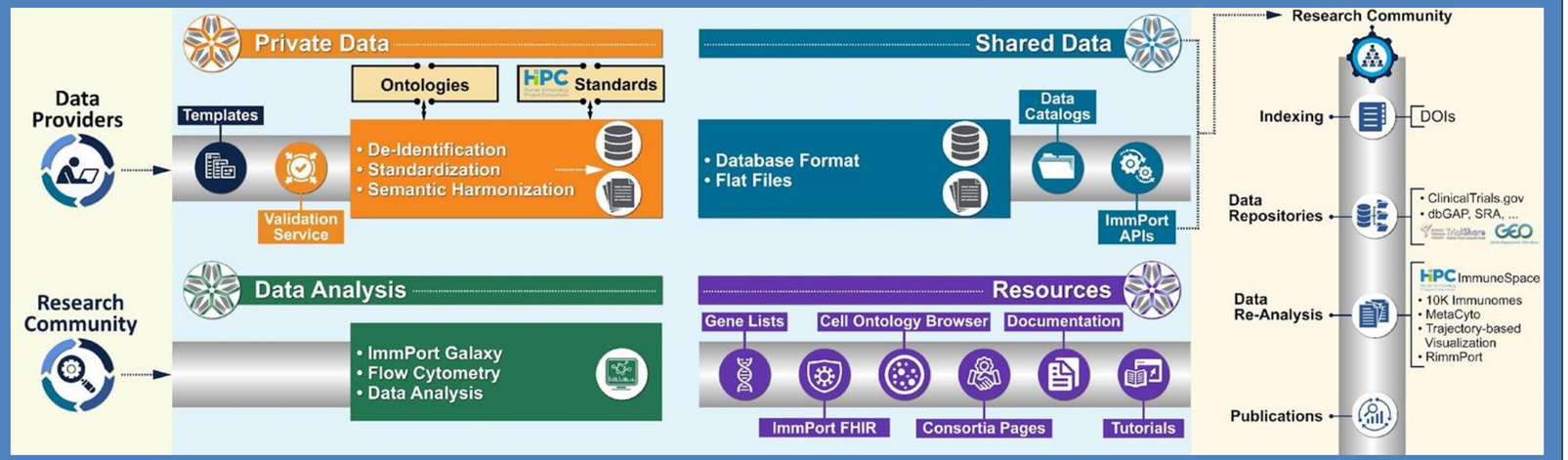
Measuring the Impact of FAIR Data: ImmPort's Role in Immunology Research

BHATTACHARYA S[1+], HYPES K[2+], SARWAL R[1], SMITH S[2], CHEPURI S[3], GU Z[2], KENDRIS E[2], KANNAN J[2], THOMSON E[2], CRAFTS M[2], and BUTTE AJ[1]

[1] University of California, San Francisco, [2] Peraton, [3] ICF, [+] Equally Contributed

Overview of the ImmPort Ecosystem

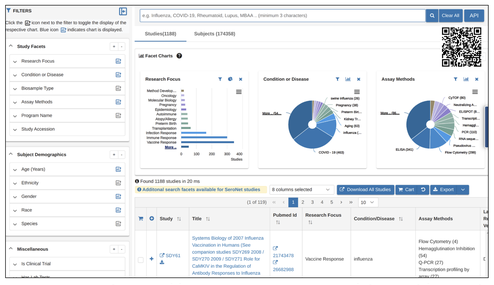


ImmPort (immport.org) is a NIAID-funded resource for sharing immunology research data and hosting web-based tools for automated analysis of data. ImmPort facilitates inter- and intra-study analysis by applying a consistent data model that captures a robust set of descriptive elements with standardized terms across publicly-shared studies. As a data sharing portal for the Division of Allergy, Immunology, and Transplantation, ImmPort focuses on studies of autoimmunity, infection and vaccine response, transplantation, and allergy.

Metadata Enabled FAIR Sharing

Cohort Discovery Tool (CDT): ImmPort provides a structured metadata collection framework and incorporates standard ontology terms to facilitate exploration and visualization of data. Advanced search of study metadata is enabled through ImmPort's CDT interface.

CDT



https://www.immport.org/shared/search

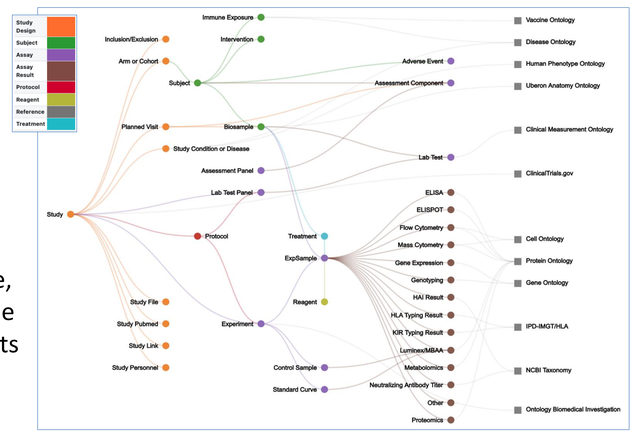
Common Data Elements (CDEs): CDEs mapped to ImmPort data elements facilitates interoperability and reuse by providing a shared set of elements relevant to disease-agnostic subject data and immunology-specific terms.

HL7 Fast Healthcare Interoperability Resources (FHIR): Translation of ImmPort data to FHIR enhances data interoperability & reuse, expands our community of users, and provides an easily accessible interface data access.

https://fhir.immport.org/

FAIR-Driven Design

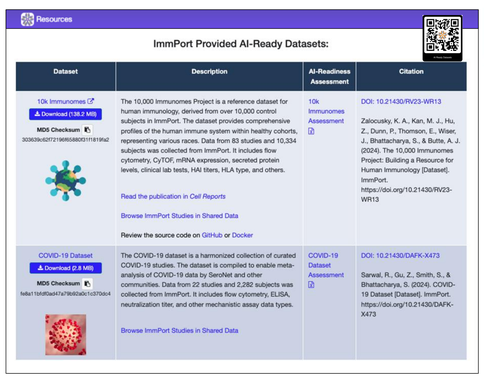
ImmPort's Data Model: Designed to collect metadata that support Findable, Accessible, Interoperable, and Reuseable (FAIR) datasets



ImmPort Data Access: ImmPort implements multiple modalities for data access and interoperability including the Global Alliance for Genomics & Health Data Repository Service Application Programming Interface (GA4GH DRS API). The GA4GH DRS API provides a unified interface for seamless data access and integration across repositories.

Enhancing FAIR with AI

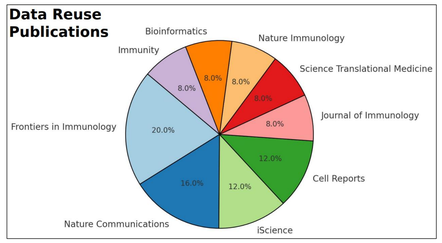
Large Language Models (LLMs): Expanding upon ImmPort's existing metadata, artificial intelligence (AI) tools are in development to extract metadata from study files. Using LLMs from OpenAI, we have developed prototypes to enrich these files with additional metadata including extracted headers and schema, generation of summaries, keywords, and more granular categorization.

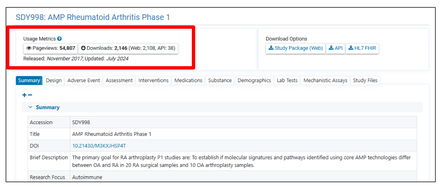


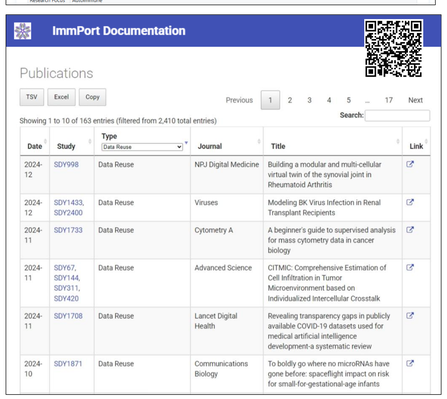
https://www.immport.org/resources/aireadydatasets

NSF National Artificial Intelligence Research Resource (NAIRR) Pilot: ImmPort prepared two AI-Ready Datasets and an AI-Readiness Assessment for the NAIRR. By associating datasets with AI-readiness metadata, we support informed selection for AI applications and pipelines

Measuring Scientific Impact







https://docs.immport.org/home/publications

Continuous Monitoring: ImmPort tracks user engagement and scientific impact metrics

Open-Source Tools: ImmPort uses open-source tools combined with code developed by the ImmPort team for metrics generation

Key Metrics: Registered user growth, dataset downloads, ImmPort resource usage, and website traffic

Publication Tracking: Displays publications referencing ImmPort data and resources

Publication Analysis: Insights from over 2,400 publications referencing ImmPort help identify dataset reuse patterns and scientific contributions

Guiding Improvements: Data-driven insights inform outreach activities and technical enhancements

Acknowledgements

The poster authors would like to acknowledge all data providers who have contributed scientific data to ImmPort, all past ImmPort team members, and the following current ImmPort team members from Peraton, UCSF, and ICF: Bhattacharya S, Campbell J, Crafts M, Chepuri S, Gu Z, Hypes K, Kannan J, Kendris E, Ketchum K, Monteiro R, Sarwal R, Smith S, Strub M, Toujas-Bernate O, Thomson E, Trilla-Flores Z, Walters B. Additionally, NIH NIAID ImmPort officers: Gururaj A, Lin D. **Funding Source:** National Institute of Allergy and Infectious Diseases, Contract # HHSN316201200036W