

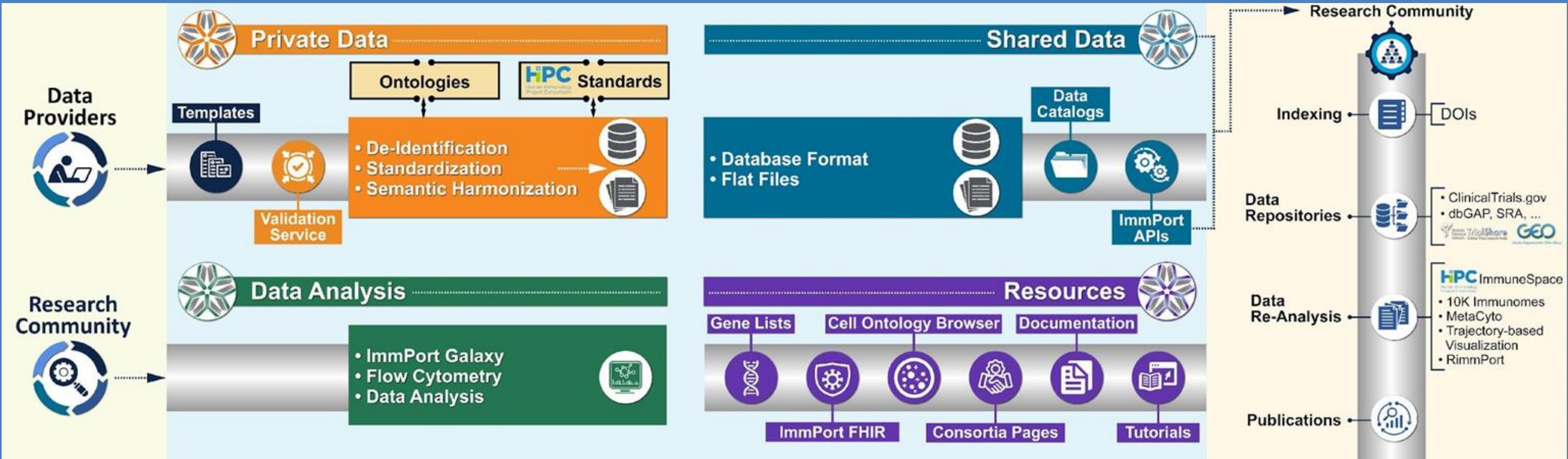
Measuring the Impact of FAIR Data: ImmPort's Role in Immunology Research

BHATTACHARYA S¹⁺, HYPES K²⁺, SARWAL R¹, SMITH S², CHEPURI S³, GU Z², KENDRIS E², KANNAN J², THOMSON E², CRAFTS M², and BUTTE AJ¹

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



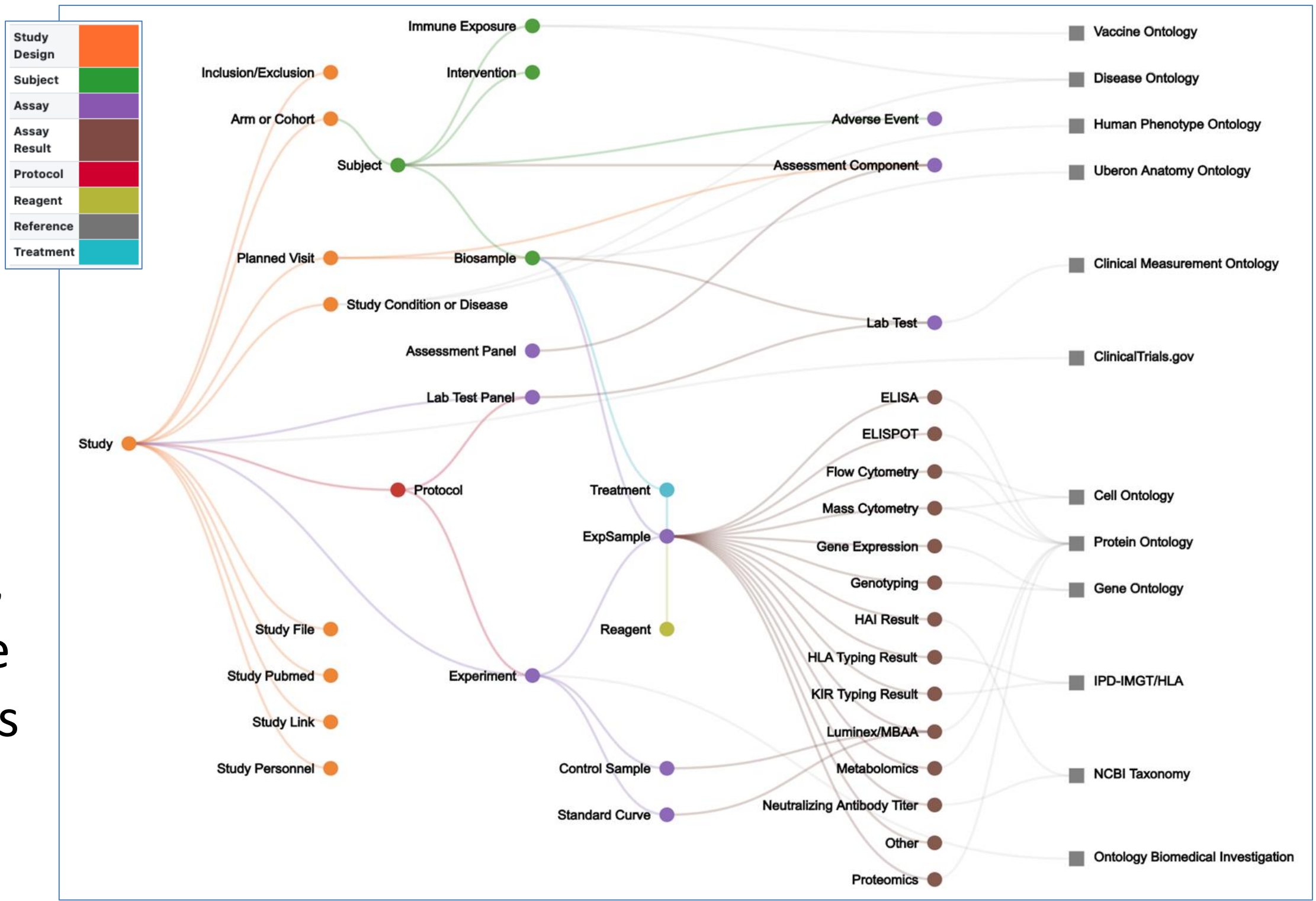
Overview of the ImmPort Ecosystem



ImmPort (import.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.

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.

ImmPort Provided AI-Ready Datasets:

Dataset	Description	AI-Readiness Assessment	Citation
10K Immunomes	The 10,000 Immunomes Project is a reference dataset for human immunology, derived from over 10,000 control subjects in ImmPort. The dataset provides comprehensive profiles of the human immune system within healthy cohorts, representing various races. Data from 83 studies and 10,334 subjects was collected from ImmPort. It includes flow cytometry, CyTOF, mRNA expression, secreted protein levels, clinical lab tests, HLA types, and others.	10K Immunomes Assessment	DOI: 10.21430/RV23-WR13
COVID-19 Dataset	The COVID-19 dataset is a harmonized collection of curated COVID-19 studies. The dataset is compiled to enable meta-analysis of COVID-19 data by Serenif and other communities. Data from 22 studies and 2,282 subjects was collected from ImmPort. It includes flow cytometry, ELISA, neutralization titer, and other mechanistic assay data types.	COVID-19 Dataset Assessment	DOI: 10.21430/DAFK-X473

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

NAIRR Pilot National Artificial Intelligence Research Resource Pilot

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

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

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.

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.

CDT

Filters: Study, Research Focus, Condition or Disease, Assay Methods, Program Name, Study Accession, Subject Demographics, Age (Year), Ethnicity, Gender, Race, Species, Microbiome

Search: [Search] [Clear]

Results: 174086

Research Focus: [Bar Chart]

Condition or Disease: [Pie Chart]

Assay Methods: [Pie Chart]

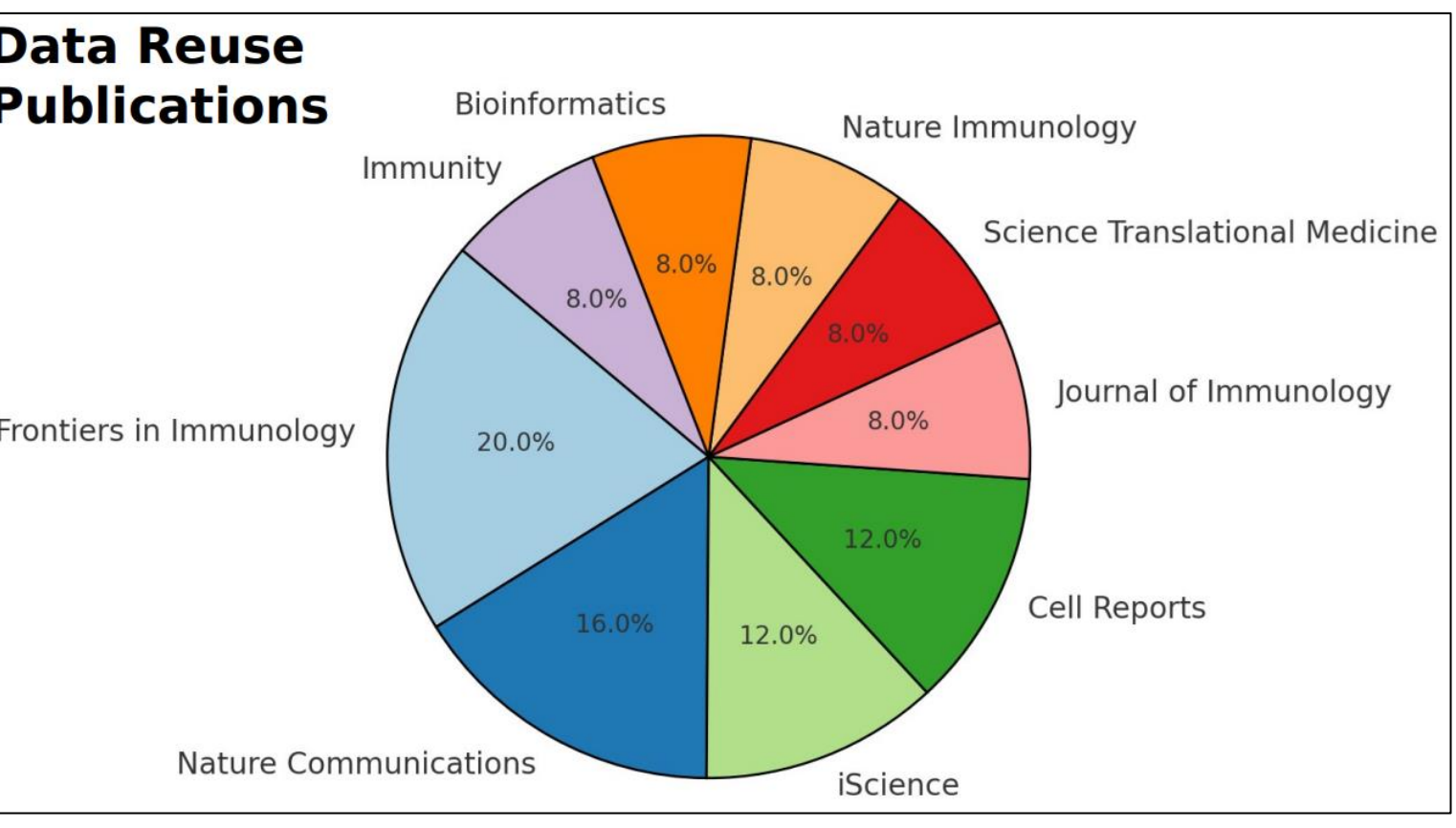
Download: [Download All Studies] [Download Selected Studies]

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

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/>

Measuring Scientific Impact



SDY998: AMP Rheumatoid Arthritis Phase 1

Usage Metrics: 4,487 Downloads, 3,148 Views, 2,108 API Calls

Download Options: [Download Package (PDF)] [API] [HL7 FHIR]

Summary: Design, Adverse Event, Assessment, Interventions, Medications, Substance, Demographics, Lab Tests, Mechanistic Assays, Study Files

Publication: SDY998

DOI: 10.21430/DAFK-X473

ImmPort Documentation						
Publications						
Date	Study	Type	Journal	Title	Link	
2024-12	SDY998	Data Reuse	NPI Digital Medicine	Building a modular and multi-cellular virtual twin of the gnosyn joint in Rheumatoid Arthritis	Link	
2024-12	SDY1433, SDY2400	Data Reuse	Viruses	Modeling BK Virus Infection in Renal Transplant Recipients	Link	
2024-11	SDY1733	Data Reuse	Cytometry A	A beginner's guide to supervised analysis for mass cytometry data in cancer biology	Link	
2024-11	SDY67, SDY144, SDY311, SDY420	Data Reuse	Advanced Science	CITMIC: Comprehensive Estimation of Cell Infiltration in Tumor Microenvironment based on Individualized Interleukin Crosstalk	Link	
2024-11	SDY1708	Data Reuse	Lancet Digital Health	Revealing transparency gaps in publicly available COVID-19 datasets used for medical artificial intelligence development: a systematic review	Link	
2024-10	SDY1871	Data Reuse	Communications Biology	To build up where no microRNA have gone before: scaffolds impact on risk for small-for-gestational-age infants	Link	

<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